A META-EVALUATION OF ESEA TITLE VII BILINGUAL EDUCATION PROJECT EVALUATIONS IN THE STATE OF FLORIDA, FISCAL YEAR 1984-85

BY

ELLA BESSIE STAES FRY

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL
FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1986

This dissertation is dedicated to

. . . the loving memory of

my father, Sidney Matthew Staes, 1889-1978,
who always wanted me to be a teacher;

and

my husband, Norman James Fry, Ph.D., 1940-1978;
who shared with me his love of the languages
and cultures of the world;

and to

. . . the celebration of life with

my mother, Ella Bessie Staes
who always believed I should be well educated and
prayed that she would live to see me reach my goal;

and

my son "Matt," James Matthew Staes Fry, age 12
who has spent his childhood on campus
patiently making all the sacrifices necessary
for me to succeed.

ACKNOWLEDGMENTS

I am deeply grateful to Dr. Madelyn Lockhart, Dean of the Graduate School, and Dr. Michael Y. Nunnery, Assistant Dean of Graduate Studies, College of Education, for their unwavering belief in me. Without their support and the support of my committee I might not have reached closure on this research project. I would like to express my appreciation to Dr. Clemens L. Hallman, chairperson of my committee, who wrote the original Title VII Teacher Trainer Fellowship proposal which brought me to the College of Education and supported my studies in bilingual-multicultural education. I also wish to express my appreciation to Dr. Sandra B. Damico, who taught me qualitative research skills and whose encouragement and assistance has made the completion of this work possible. My gratitude also goes to Dr. Robert Wright, who cared, and because he cared he taught me not to bleed when I swim with sharks. I am also most grateful to Dr. Phillip A. Clark, who challenged the students in his education leadership class to join and seek a position of leadership in a professional organization. He sponsored my nomination to Phi Delta Kappa, where I had the privilege of serving in many capacities, including chapter president and delegate. My special appreciation goes to Dr. James W. Longstreth, for his contribution to the defense of this dissertation.

I am indebted to Dr. John R. Stiefel, who has taught me to be confident and without whose support and encouragement I might not have finished. My gratitude goes to Ms. Leila Cantara, typist and nurturer, whose support and assistance was invaluable. Many loving friends provided valued assistance at the various stages in the research and provided daily doses of friendship and encouragement. My special appreciation goes to Phillip Simmons, who has been there when I needed him most, and who spent four weeks this summer entertaining my son and his surfboard while I wrote. Ms. Susan Nuttall, Ms. Chari Campbell, and Mrs. Mary Ellen Page have given invaluable assistance in revising the original draft, proofreading, keeping me in caffeine, but most importantly, being my friends.

My family has made many personal sacrifices during my graduate career. My son, Matt, and my mother, Ella Staes, have patiently waited for me to finish. Thank you.

TABLE OF CONTENTS

APPENDICES

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of
the Requirements for the Degree of Doctor of Philosophy

A META-EVALUATION OF ESEA TITLE VII BILINGUAL EDUCATION
PROJECT EVALUATIONS IN THE STATE OF FLORIDA,
FISCAL YEAR 1984-85

By

Ella Bessie Staes Fry

December, 1986

Chairman: Clemens L. Hallman
Major Department: Instruction and Curriculum

The meta-evaluator investigated the current status of evaluation in the field of ESEA Title VII funded bilingual education projects in the State of Florida for fiscal year 1984-85. A total of 12 final evaluation reports and 11 corresponding application proposal evaluation plans were analyzed. The meta-evaluation addressed the quality of the final evaluation reports not the quality of the projects. Four meta-evaluation instruments were employed in the data gathering stage. The reports were analyzed in relation to the 30 Joint Committee Standards for Educational Evaluation, while the plans were analyzed in relation to the U.S. Department of Education regulations for fiscal year 1984.

As a result of this meta-evaluation, it was concluded that contextual variables, program characteristics, and process variables were not adequately addressed in the sample reports. In addition, procedures could not be described as motivated by a thorough understanding and application of a consistent model conceptualization. The evaluation design, which provided the plan and organization for the collection of evaluation data, was the weakest area found in the meta-evaluation. No confidence can be placed in the reported results of the assessment of achievement. The sample did not meet any of the Joint Committee's 30 professional standards for good evaluation practices and only 59.4% of the applicable federal regulations. The proposals provided more descriptive information on contextual, program, and process variables and included the proposed budget, but were weakest in formulating evaluation questions, evaluation designs, comparison procedures to estimate what the performance of participants would have been in the absence of the project, data analysis procedures, and data collection methods. These same areas were also found to be inadequate in the final evaluation reports. It is recommended that Bilingual Education Act (Title VII of the Elementary and Secondary Education Act as amended in 1968) local education agency project administrators and proposal writing teams obtain technical assistance from a qualified evaluator in writing the application proposal plan. It is also recommended that the writers include specific plans for implementation evaluation and evaluation utilization in the proposals and that training be provided to enable project administrators to become more sophisticated consumers of evaluation services.

CHAPTER I
INTRODUCTION

What is the current status of evaluation in the field of bilingual education? Is it healthy and advancing methodologically or is bilingual education evaluation biased? According to Keith Baker, U.S. Department of Education, Office of Planning, Budget, and Evaluation (OPBE), "the cannons of science have been sacrificed on a wide scale to ideological needs and political expediency in bilingual research . . .'proof' of the effectiveness of bilingual instruction that vanishes into a cloud of methodological chicanery" (Baker, 1984, p. 1). Secretary of Education, William J. Bennett, in a September 26, 1985, speech before the Association For A Better New York, called bilingual education itself a failure. "After 17 years of federal involvement and after $1.7 billions of federal funds, we have no evidence that the children whom we sought to help--that the children who deserve our help have benefited" (Bennett, 1985, p. 2C). The evidence Bennett and a host of other critics across the years have been looking for should have come from nearly two decades of bilingual education evaluations and basic research in related fields. Can this lack of evidence be due to the quality of programs or the quality of the evaluations? This research assessed the merit and worth of a sample of 12 Title VII bilingual education evaluation reports from the State of Florida, fiscal year 1984-85.

Statement of Purpose

The purpose of this study was to conduct a meta-evaluation of Florida Title VII (i.e., the 1968 amendment to the Elementary and Secondary Education Act of 1965) local education agency (LEA) bilingual education project evaluations, fiscal year 1984-85. Meta-evaluation assesses the merit and worth of the final evaluation report. The focus of this study was an examination and assessment of the importance of evaluation objectives, the appropriateness of evaluation models and designs, the adequacy of implementation of the designs, the technical adequacy of data analysis, and the quality and importance of evaluation results. The researcher performed an indepth analysis using criteria related to evaluation model characteristics, evaluation design characteristics, data collection procedures, testing instruments, data analysis techniques, and general evaluation methodologies, as well as discrepancies between proposed evaluation plans and summative evaluation reports in order to answer the following four questions:

1. Which context and process variables are addressed in Title VII project summative evaluation reports within the State of Florida?

2. What are the characteristics of Florida Title VII evaluation models, designs, and reports in terms of information coverage, content, and procedures?

3. Do Florida Title VII evaluation reports meet the current accepted professional standards of good evaluation practices, as detailed by the Joint Committee on Standards for Educational Evaluation (1981), _Standards for Evaluations of Educational Programs, Projects, and Materials_?

4. Are there major discrepancies between LEA evaluation proposals and summative evaluation reports? What is the nature of these discrepancies?

Significance of the Study

> If evaluations are to provide proper guidance, the evaluations themselves must be sound. Among other considerations, they must be focused on the right questions, accurate in their portrayals, free from bias, understandable, and fair to those whose work is under examination. (Stufflebeam & Shinkfield, 1985, p. 183)

This meta-evaluation of Title VII of the Elementary and Secondary Education Act (ESEA) proposals and final evaluation reports adds to the knowledge base of bilingual education. Tallmadge, Lam, and Camarena (1985a) warned that

> all relevant characteristics of students, settings, and treatments must be carefully documented as an integral part of any bilingual education program. Failure to do so would run the risk that educationally significant relationships would be obscured whenever data were pooled across different types of students, treatments, and/or settings. (p. xii)

Tallmadge et al. questioned the effectiveness of the variations in educational treatments which all go under the name of bilingual education for the various types of students in the multiplicity of settings. Evaluators cannot answer that question until they have documented the range of treatments, students, and settings. Instruments used in this meta-evaluation research project provided for the collection of such contextually rich information. In addition, this research adds to the knowledge base of bilingual education practices and their effects at the LEA level by accurately describing treatment as implemented rather than treatment as intended in the proposal. Tallmadge et al. (1985a) indicated that "the actual treatment, unfortunately, may bear little resemblance to what was intended and may, consequently, have a very low construct validity relative to what the study set out to evaluate" (p. xiii).

This meta-evaluation also serves as an aide to designers of future bilingual programs and the planners of any national data reporting systems. Fuentes' (1986) contended that "a need thus exists for empirically grounded information that can guide planners in designing the most effective programs. For example, educators need information on the effectiveness of ongoing practices and on the need for improvement of practices" (p. 1). This meta-evaluation research satisfies Fuentes (1986) requirement and fits into Sanders' (1981) first stage for the theoretical design of a national system for monitoring federally-funded bilingual programs. Sanders' (1981) conceptual analysis stage is composed of case study data and survey data from Title VII projects "to discover the extent to which certain project characteristics exist and to identify further distinctions that must be incorporated into the reporting system" (p. 10). Stage two is the pilot stage. Once the pilot reporting program is field tested, then stage three involves a full-scale implementation of the design and continuous monitoring of operation for quality control. State meta-evaluations conducted in states which have a high concentration of Title VII programs for limited-English-speaking students would provide for the United States Department of Education, Office of Bilingual Education and Minority Language Affairs (OBEMLA) a much more accurate regional and national picture of the current status of LEA evaluation practices than is currently available from past studies and would contribute greatly to OBEMLA's current efforts at systematizing bilingual education program evaluation.

As further indication of the willingness of the federal government to improve the quality of bilingual education evaluations, the 1984 bilingual reauthorization bill provided for the establishment of two bilingual education evaluation assistance centers. At the present time, only one center is currently in operation, located at Georgetown University. Dr. Michael J. O'Malley, Director of the Title VII Evaluation Assistance Center, funded by OBEMLA, expressed great interest in this meta-evaluation of Florida Title VII project evaluations as the results, in his opinion, would provide timely contextual information useful in the designing and implementation of an evaluation system for bilingual education (M. J. O'Malley, personal communication, May 15, 1986).

The results of this meta-evaluation study provide a description of current evaluation practices within Florida, data deemed valuable by the Miami Bilingual Education South East Service Center (BESES), the bilingual consultants in the Florida Department of Education, the Bilingual Education Evaluation Service Center at Georgetown University, and Office of Planning, Budget, and Evaluation, U.S. Department of Education.

Scope of Study

This study was confined to federally funded Title VII LEA bilingual project proposals and summative evaluations within the State of Florida. Locally funded bilingual projects are not required to evaluate effectiveness due to the absence of bilingual education legislation within the State of Florida. The study was further confined to the fiscal year 1984-85, since final evaluation reports were not required to be submitted to OBEMLA until the December following the close of each grant (fiscal) year, making 1984-85 the most recent available data.

Limitations

Although final evaluation reports come under the "government in the sunshine" regulations within the State of Florida, making them theoretically accessible to the public, the documents used in this study were limited to those proposals and final evaluation reports voluntarily submitted by districts which received federal funds for Title VII projects in the fiscal year 1984-85. Also, Florida does not have legislation which authorizes or funds bilingual education programs. Limitations must be placed on the generalizability of the reported results to other states, especially states which do have bilingual education legislation.

The evaluator, unlike the researcher in his or her laboratory, must function in the context of local schools where there is a multiplicity of variables which must be identified and described but are beyond the authority of the evaluator to control. Limitations of context and degree of implementation must be placed on the generalizability of the reported results.

Definition of Terms

Definitions of the following terms are presented in order to clarify the meaning of the study and to avoid needless ambiguity. Contextually relevant terms are presented together rather than by alphabetical order. Where applicable, the author or source for the definition is identified.

1. Bilingual education. Bilingual education provides special programs for non- and limited-English proficient language minority students which have the dual objectives of developing proficiency in English language skills while preventing the student from falling behind English proficient peers in content areas (Tallmadge, Lam,& Camarena, 1985b, p. vii).

A. Title VII. The 1968 Bilingual Education Act authorized bilingual education by amending Title VII of the Elementary and Secondary Education Act of 1965. Title VII, therefore, is used to refer to bilingual education projects.

B. LEP. Limited English Proficient (LEP), is a category defined by predetermined cutoff score on a language assessment test. This category may also include non-English-speaker (NES) and limited-English speaker (LES).

> The term "limited English proficiency" when used with individuals means (a) individuals who were not born in the United States or whose native language is a language other than English, (b) individuals who come from environments where a language other than English is dominant, as further defined by the Commissioner by regulation, and individuals who are American Indian and Alaskan Native students and who come from environments

> where a language other than English has had a significant impact on their level of English language proficiency, subject to such regulations as the Commissioner determines to be necessary; and, by reason thereof, have sufficient difficulty speaking, reading, writing, or understanding the English language to deny such individuals the opportunity to learn successfully in classrooms where the language of instruction is English. (United States Department of Education, 1983, p. E2)

C. ESL. English as a Second Language (ESL) is a component of the bilingual program that teaches English to speakers of other languages in an intensified manner, taking into consideration the challenges of each student within the context of their primary language and their English language skills (Martin, 1981, p. 21).

D. Lau categories. A system for categorizing students by language proficiency, the Lau categories grew out of the Lau v. Nichols Supreme Court Decision and subsequent Lau Remedies. The Lau categories are (A) exclusively monolingual speaker of a language other than English, (B) predominantly speaker of a language other than English although speaks some English, (C) equally speaker of a language other than English and English, (D) predominantly speaker of English but speaks some of the other language as well, and (E) exclusively monolingual speaker of English. Students who are classified Lau-A and Lau-B are considered to be limited English proficient (LEP).

E. LEA. A local education agency (LEA) is a school district.

F. OBEMLA. OBEMLA is an acronym for the Office of Bilingual Education and Minority Language Affairs, U.S. Department of Education.

G. BESES. BESES is the Bilingual Education South East Service Center.

2. Evaluation. Evaluation is the assessment of worth and merit (Lincoln & Guba, 1984); a process of delineating, obtaining, and providing useful information for judging decision alternatives (Stufflebeam, 1978); and "a type of disciplined inquiry undertaken to determine the value (merit and/or worth) of some entity . . . the evaluand, such as a treatment, program, facility, performance, and the like . . . in order to improve or refine the evaluand (formative evaluation) or to assess its impact (summative evaluation)" (Lincoln & Guba, 1984, p. 9).

A. Meta-evaluation. Meta-evaluation is defined as the evaluation of evaluations (Scriven, 1969); the assessment of the worth and merit of an evaluation, including all levels of evaluation above the primary evaluation (Stufflebeam, 1978); the process of delineating, obtaining, and using descriptive and judgmental information about the utility, practicality, ethics, and technical adequacy of an evaluation in order to guide the evaluation (formative meta-evaluation) and to report publicly its strengths and weaknesses (summative meta-evaluation) (Stufflebeam, 1981a).

1.) Primary evaluation. A primary evaluation is the assessment of worth and merit of some entity, such as an ESEA Title VII education project, and is an evaluation that is the subject of a meta-evaluation (Stufflebeam, 1974a).

2.) Formative meta-evaluation. A formative meta-evaluation is a meta-evaluation that is conducted at the same time as the primary evaluation and is intended to guide the primary evaluation (Stufflebeam, 1981a).

3.) *Summative meta-evaluation*. Summative meta-evaluation is a study that judges the worth and merit of completed evaluations (Stufflebeam, 1981a), such as the meta-evaluation of the sample of 12 Florida ESEA Title VII bilingual education final evaluation reports for the fiscal year 1984-85.

4.) *Standards*. Widely shared principles for the measure of worth and merit of an evaluation defines the word standards.

5.) *Guidelines*. Guidelines are procedural suggestions intended to help evaluators meet evaluation standards (Stufflebeam, 1981a).

6.) *Pitfalls*. Mistakes that are commonly made by persons who are unaware of the importance of a given principle of sound evaluation are pitfalls (Stufflebeam, 1981a).

7.) *QUEMAC*. QUEMAC is a series of questions which, when answered, reveals the logical structure of evaluation studies (question, event or object, method, answer, concept) (Gowin & Millman, 1978).

B. *Discrepancy Evaluation Model*. The Discrepancy Evaluation Model (DEM), developed by Provus (1972), is a comparison between the present state of affairs with "what should be;" a matter of making judgments about the worth or adequacy of an object based upon the discrepancy between standard and performance measure (Brannon, 1985, p. 5).

1.) *(S) Standard*. A standard is a list, description, or representation of qualities or characteristics the object should possess--the ideal--how something "should be" (Steinmetz, 1983, p. 2).

2.) *(P) Performance measure*. A performance measure is the actual characteristics of the object to be evaluated (Steinmetz, 1983, p. 2).

C. Impact evaluation. Impact evaluation, typically quantitative in approach, is used to assess a social program's overall performance, accomplishments, and the effects of the program on participant outcome. Impact evaluation asks whether the program is a success or failure. As used by the General Accounting Office (GAO), impact evaluation provides Congress with an assessment of the performance of a federally-funded social program, whether a social program is achieving its legislative objective, and how the program is affecting the intended beneficiaries.

3. Research. For the purposes of this study, research is defined as a type of disciplined inquiry undertaken to resolve some problem in order to achieve understanding or to facilitate action--problem resolution or amelioration (Lincoln & Guba, 1984).

Overview of Chapters II, III, IV, and V

An introduction to the meta-evaluation was provided in Chapter I. The purpose, significance, scope, and limitations of the study were identified and relevant terms were defined. Literature related to the historical context for the evaluation of bilingual education projects, evaluation as a profession, and meta-evaluation is reviewed in Chapter II, along with applicable literature on the selection of meta-evaluation criteria and appropriate standards, the application of meta-evaluation theory to the field of bilingual education, and program evaluation versus evaluation research. In Chapter III, the methodology of the meta-evaluation is discussed. The results of the meta-evaluation are presented in Chapter IV. The results are discussed and summarized in the order of the four main research questions. In Chapter V, the final chapter of the report, the meta-evaluation findings are discussed, implications are identified, and recommendations are offered.

CHAPTER II
REVIEW OF RELATED LITERATURE

This review of related literature places the fields of bilingual education, evaluation, and the development of meta-evaluation in historical context. Brief introductory segments introduce the legislative history of bilingual education and the emergence of evaluation as a profession, followed by a discussion of the development of the theory of meta-evaluation. A brief discussion of studies which have applied meta-evaluation methodology to bilingual education program evaluation follows. Questions of the appropriateness of using local LEA program evaluation data for the purpose of reporting research on the effectiveness of Title VII bilingual education programs to Congress are discussed. The chapter concludes with a discussion of the distinction between evaluation research and program evaluation.

Historical Context for Evaluation in Bilingual Education

The federal government first became involved with the problems of non-native speakers of English with the passage of the Civil Rights Act of 1964. This was closely followed by the Bilingual Education Act of 1968, otherwise known as the Title VII amendment to the Elementary and Secondary Education Act of 1965, which provided supplemental funding for school districts to establish programs to meet the special educational needs of large numbers of children of limited English speaking ability from low income families in the United States. Neither the Civil Rights Act nor the ESEA Title VII amendment specified what actions needed to be taken to assure language minority students equal educational opportunities. In January of 1974, the U.S. Supreme Court affirmed that school districts would be compelled under the Title VI of the Civil Rights Act of 1964 to provide special language programs for children who speak little or no English to provide them with an equal educational opportunity. It was this supreme court decision which spurred most state and local educational agencies to design and implement bilingual educational programs.

Since the Bilingual Education Act of 1968 did not specifically require program evaluations, little was known about success of programs or student achievement for nearly 10 years. Development Associates (1973), under contract to the U.S. Office of Education, submitted the first part of a process evaluation of Title VII programs in December of 1973. The emphasis of that evaluation was on the extent of adherence to guidelines by individual Title VII projects and the relationship between such adherence and success based on subjective ratings by evaluation team leaders. The question of assessment of student achievement attributable to Title VII program attendance was not addressed.

The Bilingual Education Act of 1974 removed the criterion of low income and added, among other things, the requirement that the U.S. Commissioner of Education and the National Advisory Council for Bilingual Education would report to the U.S. Congress on the state of education in the nation, including an evaluation of Title VII activities. According to the United States Commission on Civil Rights (1975), the nature of evaluation was still not clear and support was limited for the overall program.

It was not until 1976 that specific "Title VII Regulations" for LEA program evaluations were published in the Federal Register. Tallmadge, Lam, and Camarena (1985a) commented that guidance provided for completing the required evaluation was minimal. They posited that three factors--lack of evaluation expertise at the local level, low priority and low funding levels, and technical difficulties inherent in conducting bilingual program evalautions--"combined to produce the not surprising outcome of basically useless data" (p. viii).

The U.S. Department of Education, Office of Bilingual Education and Minority Language Affairs (OBEMLA), is currently in the process of exploring ways to improve the quality of bilingual education evaluations. A three-year contract was granted to SRA Technologies in 1985 to develop and field test an evaluation system for bilingual education that would incorporate methdologically sound designs and procedures. A major goal for this project is that it be useful for the improvement of program at the local level and yield outcome data appropriate for comparisons and aggregation at the federal level (Tallmadge, Lam, & Camarena, 1985b).

Historical Context: Evaluation as a Profession

The 1964 federal mandate for evaluation of ESEA Title I federally funded compensatory education programs caught educators and academicians alike unprepared. College professors from several disciplines including psychology and education responded to the needs of the local school districts for evaluation consultants to help them meet the new federal evaluation requirements. James Popham (1981), UCLA Center for the Study of Evaluation, reminisced "most of us were retreads from other fields, typically research and measurement, since it was thought in those days that educational evaluation was little more than warmed over research with applied focus" (p. 30).

The growing need for consultants with special technical training in evaluation methodology gave rise to training institutes and professional organizations which provided much needed attention to the skills of the field and the development of theory. By the late 1970s and early 1980s evaluation had developed into a separate nascent profession with universities providing degree programs, the establishment of evaluation societies, the publication of evaluation journals, and the development of at least two sets of professional standards by which the profession could begin self-regulation and improve the quality of its procedures and products. When asked in May of 1980 to comment on the status of educational evaluation, Nick L. Smith, Director, Research on Evaluation, Northwest Regional Laboratory, commented,

> We haven't been at this evaluation business very long, and we are not terribly sure whether we are adroitly

> navigating each new turn or merely trying to keep upright through the changing economic currents, the encroaching banks of legislation and judicial control, and the shoals of special interest groups. We have had to gain our balance quickly, having been launched in the swift currents of the ESEA legislation in the mid-sixties. (Smith, 1981c, p. 19)

With this growing self-consciousness of emerging into an independent profession, theory and methodology worked together to provide a way "to assure quality evaluation services, to guard against or deal with malpractice or services not in the public interest, to provide direction for improvement of the profession and to promote increased understanding of the evaluation enterprise" (Stufflebeam & Shinkfield, 1985, p. 34).

Historical Context: Meta-Evaluation

Although Scriven (1969) is given the credit for introducing the concept of meta-evaluation and defining it as "the evaluation of evaluation," it was Pedro T. Orata who introduced the concept. Orata published an article entitled "Evaluating Evaluation" in the May, 1940 issue of Journal of Educational Research, in which he critiqued Tyler and Wrightstone's new evaluation practices in the 30 schools of the Eight Year Study, versus the traditional testing and measurement practices of the day. Orata recommended "for the evaluation practices to be evaluated and subsequently modified so as to provide, in fact as well as theory, new tests for new needs" (p. 653).

Michael Scriven (1969), in an article entitled "An Introduction to Meta-Evaluation," gave the name meta-evaluation to the process of evaluation of evaluation, and described the process both theoretically as "the methodological assessment of the role of evaluation" and practically as "concern with the evaluation of specific evaluation performances" (p. 36). Scriven (1976) applied the underlying concept in a brief review on the assessment of a specific educational product--the social science curricula.

Scriven (1969) did not offer an elaborate conceptualization of meta-evaluation or discuss the theoretical aspects of meta-evaluation, nor did he work out a logical structure of explicit methodology. It was Daniel Stufflebeam who expanded the theory and set up a detailed model. "Good evaluation requires that evaluation efforts themselves be evaluated . . . it is necessary to check evaluations for problems such as bias, technical error, administrative difficulties, and misuse" (Stufflebeam, 1974a, p. 1). "If all evaluations are potentially faulty, then, theoretically, each one requires scrutiny and assistance from meta-evaluation" (Stufflebeam, 1981a, p. 153).

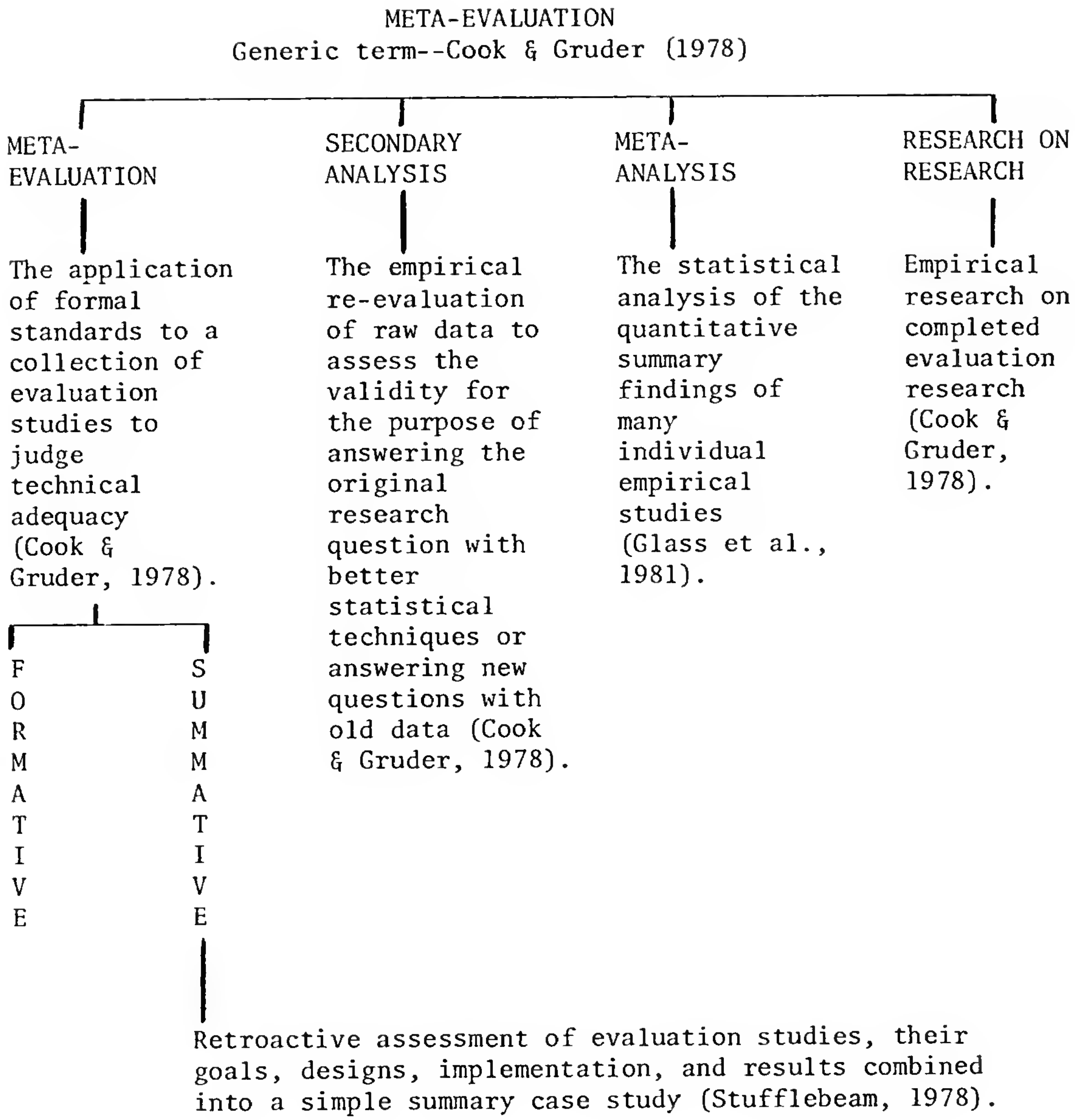
Stufflebeam (1974a), in a 106-page monograph, addressed the conceptual and practical development of meta-evaluation. Part I of the monograph detailed a rationale-justification for meta-evaluation, suggested what criteria should be used to guide the development of methodology, and outlined six classes of problem areas that jeopardize evaluation which need to be addressed by meta-evaluation and presented a logical structure for designing meta-evaluation studies. Part III outlined four formative and one summative meta-evaluation designs, giving substance to the conceptualization.

Cook and Gruder (1978) described four attempts at the employment of meta-evaluation theory and methods and developed a classification of seven meta-evaluation research models for improvement of technical quality and relevance of empirical summative evaluations. Cook and Gruder conceived of meta-evaluation as the genus and unifying theory under which three species of research traditions functioned: the application of formal standards to a collection of evaluation studies to judge technical adequacy; the empirical reevaluation of raw data to assess the validity for the purpose of answering the original research question with better statistical techniques, or answering new questions with old data--secondary analysis; and research on research. Gene Glass (Glass, McGaw, & Smith, 1981) added one additional species of research tradition, meta-analysis, one approach to research integration which Glass defined as the statistical analysis of the quantitative summary findings of many individual empirical studies.

This meta-evaluation research project employed Stufflebeam's (1978) fifth meta-evaluation design, a retroactive assessment of evaluation studies--their goals, designs, implementation, and results combined into a simple summary case study (see Figure 1).

Selection of Meta-Evaluation Criteria

Stufflebeam stated that "good evaluation requires that evaluation efforts themselves be evaluated" (1974b,p. 1). Meta-evaluation sums up the overall merit of an evaluation by assessing the extent that an evaluation is technically adequate, useful, ethical, and practical (Stufflebeam, 1978, p. 22).



META-EVALUATION
Generic term--Cook & Gruder (1978)

META-EVALUATION

The application of formal standards to a collection of evaluation studies to judge technical adequacy (Cook & Gruder, 1978).

FORMATIVE

SUMMATIVE

Retroactive assessment of evaluation studies, their goals, designs, implementation, and results combined into a simple summary case study (Stufflebeam, 1978).

SECONDARY ANALYSIS

The empirical re-evaluation of raw data to assess the validity for the purpose of answering the original research question with better statistical techniques or answering new questions with old data (Cook & Gruder, 1978).

META-ANALYSIS

The statistical analysis of the quantitative summary findings of many individual empirical studies (Glass et al., 1981).

RESEARCH ON RESEARCH

Empirical research on completed evaluation research (Cook & Gruder, 1978).

Figure 1. Meta-evaluation design.

Of major concern to the meta-evaluator is the selection of appropriate criteria for judging the meta-evaluation. Scriven (1969) proposed technical adequacy and Stufflebeam (1978) added utility and cost effectiveness. The Phi Delta Kappa Study Committee listed 11 specific criteria for judging technical adequacy, utility, and cost effectiveness (Stufflebeam, 1974a; Stufflebeam & Madaus, 1983). The four criteria of technical adequacy are (a) internal validity, (b) external validity, (c) reliability, and (d) objectivity. The seven standards of utility are (a) relevance, (b) importance, (c) scope, (d) credibility, (e) timeliness, (f) pervasiveness (dissemination to intended audiences), and (g) cost effectiveness.

Standards, Guidelines, and Pitfalls

The meta-evaluator must deal with the question of which guidelines and pitfalls are most influential in meeting given standards of sound evaluation. Stufflebeam (1981a) explained that,

> By defining standards one is essentially developing a set of dependent variables regarding the outcomes of evaluation studies. By defining guidelines and pitfalls one is essentially developing a set of independent variables for assessing evaluation operations. Once defined, such sets of independent and dependent variables would be a good resource for deriving and testing hypotheses about how an evaluation's performance, in meeting the procedural guidelines and avoiding common pitfalls, influences the evaluation's overall satisfaction of the standards. Such guidelines and standards constitute a basis for the actual design and conduct of meta-evaluation studies. (p. 153)

Daniel Stufflebeam and 200 other people worked for four years to develop professional standards for educational evaluation. The project was sponsored by 12 professional organizations who appointed

17 members to act as the Joint Committee on Standards for Educational Evaluation (Stufflebeam, 1981b). As director of the project, Stufflebeam affirmed that "one characteristic of a profession is the maintenance of high standards for achievement and ethical conduct" (p. 40). The Joint Committee developed and published 30 standards of good practice presented in four groups that correspond to the four main concerns about evaluation--its utility, feasibility, propriety, and accuracy (Joint Committee on Standards for Educational Evaluation, 1981). Together, the 30 standards are a working philosophy of evaluation, defining principles that should guide and govern evaluation efforts, including meta-evaluation efforts, and offer practical suggestions for observing these principles. Stufflebeam (1981b) recommended the use of the standards for all stages of evaluation including meta-evaluation. "Now that the field has articulated 30 standards of good practice, it will be important to ascertain the quality of evaluations in relation to each of the standards" (p. 43).

The Joint Committee is not the only source of standards. The Government Accounting Office (GAO) standards were designed to guide in the judgment of impact evaluations. Finding that the Joint Committee's evaluation standards were best applicable to education and the GAO's standards were only applicable to impact evaluations, the Evaluation Research Society (now called the American Evaluation Association), whose membership's diverse evaluation interests span a broad spectrum of fields, initiated their own ERS Standards Committee. The end product of this committee's work was the

Evaluation Research Society Standards for Program Evaluation adopted by the Evaluation Research Council (Evaluation Research Society, 1984).

The meta-evaluator must have access to accurate information on the numerous testing instruments used in all of the evaluations he or she encounters. The Mental Measurement Yearbooks (Buros, 1972; Mitchell, 1985) might well be called the standard reference for tests in use. Campbell and Stanley's (1963) Experimental and Quasi-Experimental Designs for Research is still considered the standard for questions concerning the evaluation of alternative experimental designs.

Meta-Evaluation Instruments, Techniques, and Devices

The literature is surprisingly full of checklists for organizing meta-evaluation. Scriven's (1974, 1985) Key Evaluation Checklist is a synthesis of his contemplation of meta-evaluation. He has at times referred to his checklist as the multimodel of evaluation, reflecting his view that evaluation involves multiple dimensions, should employ multiple perspectives, involves multiple levels of measurement, and must employ multiple methods. There are 18 items in the Key Evaluation Checklist, the 18th referring to meta-evaluation.

> The evaluation must be evaluated, preferably prior to (a) implementation, (b) final dissemination of report. External evaluation is desirable, but first the primary evaluator should apply the Key Evaluation Checklist to the evaluation itself. Results of the meta-evaluation should be used formatively but may also be incorporated in the report or otherwise conveyed (summatively) to the client and other appropriate audiences. (Scriven, 1985, p. 1)

Scriven perceived evaluation as a data-reduction process where the early stages help characterize a program or product and the later stages help to assess its validity.

Stufflebeam's (1974a) monograph, Meta-Evaluation, is organized like a giant 106-page outline, the whole of which could be reduced to a brief checklist for doing a meta-evaluation. If not the whole document, then the discussion of "Problems that Jeopardize an Evaluation" could certainly be used as such. There is one checklist in the monograph, "An Administrative Checklist for Reviewing Evaluation Plans," designed for formative meta-evaluation but quite usable as a supplement to a summative meta-evaluation.

Millman's (1981) "The Checklist" focuses on the merit or worth of the evaluator's assessment of the program or product and the merit or worth of the evaluation itself. Questions are directed to three components of the program or product and the evaluation: (a) the preconditions or preliminaries to the conduct of the program or development of the product and to the evaluation, (b) the effects of the program or product and the evaluation, and (c) the utility of the program or product and the evaluation.

There are a number of forms a meta-evaluation can take. Smith (1981b) suggested the following as possible topics of inquiry for a meta-evaluation: design, management, instruments, data, results, impact, personnel, purpose, setting, reporting, the use of formal criticism, or any combination of the above. Stevenson, Longabaugh, and McNeill (1979), in considering meta-evaluation practices in the field of human services, have developed a model for purposive evaluation applying a grid of extent of planned decision relevance and intended locus of impact, where within each cell the variety of purposes for evaluation can be classified "to determine the extent to which evaluators and clients intended the study to fulfill particular purposes, and the extent to which their expectations were met" (Stevens et al., 1979, p. 46). To date, little has been done in educational evaluation with the question of overt and covert intentions for decision making.

Philosophical analysis was proposed by Gowin (1981) who added QUEMAC Value Appraisal to the list of forms which meta-evaluation may take. QUEMAC, a construct for detecting the logical structure of an evaluation, is an acronym for questions, events or objects, methods, answers, and concepts (Gowin, 1981, p. 300). The meta-evaluator, following the QUEMAC guide, answers six questions about the evaluation study. In answering these questions he or she is able to uncover embedded value concepts. Gowin advised that "value theory can help in evaluation appraisal, and when applied critically to evaluation practice, should show us how we treat value questions and also to exhibit what our values are" (Gowin & Millman, 1978, p. 9).

Smith (1981a) suggested that the meta-evaluator illuminate the nature of an evaluation while simultaneously assessing its quality. He recommended the use of the techniques of formal criticism--description, analysis, interpretation, and evaluation. The nature of an evaluation may be illuminated just as adequately through the application of qualitative methodology.

Discrepancy Analysis Techniques

Tallmadge, Lam, and Camarena (1985a) have observed that "it is treatment as implemented, not the treatment as intended, that is evaluated. The actual treatment, unfortunately, may bear little resemblance to what was intended and may, consequently, have very low construct validity relative to what the study set out to evaluate" (p. xiii). Sanders (1981) continued along this line by observing,

> experiences with local projects have shown, however, that the actual objectives and consequences of a project tend to drift away from stated objectives as the project evolves. What is written and what is done differ and the design for monitoring Title VII bilingual education projects should consider the latter. (p. 5)

One approach to analyzing the discrepancies between what is proposed and what is done as chronicled in the final evaluation report is to apply an adaptation of the Discrepancy Evaluation Model (DEM) proposed by Provus (1972). A modification of Provus's Stage II places the proposal's evaluation plan as the standard (S), the representation of how the evaluation should be. The final evaluation report represents the performance measure (P), the actual evaluation as completed and reported. Comparison of the congruence of (S) against (P) yields discrepancy information (D). Steinmetz (1983) concluded that discrepancy analysis is a matter of making judgments about the worth or adequacy of an object based upon (D) information which is the result of a comparison of (S) against (P). The model leaves the questions of criteria used and tolerable quantity of discrepancy to the judgment of the evaluator who must answer the following three questions:

1. Is the information on each program element complete?

2. Is the information reliable and valid?

3. Are the discrepancies uncovered ones which will significantly diminish the program's chances of success?

Application of Meta-Evaluation Theory to Bilingual Education

Currently, the best methodology available for assaying the soundness and technical adequacy of evaluations is meta-evaluation. Still a relatively new procedure, each meta-evaluation that is completed adds to and extends what is known about meta-evaluation theory and methodology. Gowin and Millman (1978) considered this a "fertile area for research in evaluation" (P. 6). Stufflebeam (1981a) asserted that meta-evaluation can play a vital role in advancing evaluation as a profession by providing quality assurance, self-regulation, renewal, and accountability. "It is strongly recommended that educational evaluators make meta-evaluation a matter of common practice in their work" (p. 158).

Lois-Ellen Datta (1981a), of the National Institute of Education, provided evidence for the need for meta-evaluation when she said that "the effectiveness of evaluation has been questioned since at least 1969 when Guba, followed by Wholey et al. (1970) and Weiss (1972) asserted that most federally-funded evaluations had little impact on programs, practices or policy" (p. 125). Looking specifically at reviews of the literature on the effectiveness of bilingual education evaluations (Alkin, Kosecoff, Fitz-Gibbon, & Seligman, 1974; Baca, 1984; Baker & de Kanter, 1981; Burry, 1979; Dulay & Burt, 1979a, 1979b; English, 1983; Martin, 1981, 1982a, 1982b; Okata, 1983; Zappert & Cruz, 1977), the overwhelming consensus indicated the continued presence of serious technical-methodological flaws, which seem to go uncorrected with time and without having any deleterious effect on the funding of continuation grants.

Keith Baker (1984), Office of Planning, Budget, and Evaluation, U.S. Department of Education, was one of the more outspoken critics of bilingual education evaluation practices. In a paper presented at the American Educational Research Association (AERA) convention, Baker (1984) offered the following indictment.

> We have come to the conclusion after extensive study of the research literature in bilingual education that this discipline is pervaded by the subordination of the cannons of scientific research to ideological and political needs. . . . This paper briefly reviews some selected examples of bilingual education research and their methodological faults. . . . The thesis explored is that the extensive methodological problems found in this literature are so severe and so strongly biased toward conclusions supporting the bilingual ideology that they cannot be accounted for as merely examples of incompetence. The only explanation that can account for the observed flaws in this literature is that there has been widespread abandonment of proper research procedures to ensure the dominance of a particular ideology in the education of minority language children. (p. 1)

Evaluation criticisms can neither be validated nor refuted at the local LEA level within the State of Florida until there is a thorough evaluation of the most recent bilingual evaluation reports, hence the need for this meta-evaluation. Stufflebeam and Shinkfield (1985) declared that "evaluators have a professional obligation to ensure that their proposed or completed evaluations are subjected to competent evaluation" (p. 321). Stufflebeam (1981a) exhorted, "people and agencies throughout educational evaluation need to sustain and increase their efforts to develop and use meta-evaluation" (p. 158).

One of the first meta-evaluations of Title VII project evaluations was done by Alkin, Kosekoff, Fitz-Gibbon, and Selgiman (1974), at the UCLA Center for Study of Evaluation (CSE). The authors reported findings on a nationwide sample of 42 LEA Title VII bilingual projects to determine the impact of evaluations upon decision-making at the local and federal levels. CSE raters could not make any judgments about the effectiveness of the projects based on the final evaluation reports. Alkin et al. reported that they felt confident that, in many cases, formative evaluation was conducted by evaluators and found useful by project directors, but was not evident in evaluation reports. When asked directly, the majority of project directors indicated that evaluation had been influential in decision making, particularly on those decisions concerning modifications for the following school year. Although refunding decisions were made before final evaluation reports were received by federal personnel, CSE staff found that projects in urban areas were large, had high funding levels, employed specialists as evaluators who produced sophisticated and comprehensive evaluation reports, which tended to favorably impress federal monitors, and thus had the best chance of influencing federal decision makers.

The next year, Campeau, Roberts, Bowers, Austin, and Roberts (1975) examined 175 final evaluation reports from bilingual programs but judged that only 8 of them merited site visits. Zappert and Cruz (1977) looked at 108 project evaluations and 76 research studies to judge the effectiveness of bilingual education on student performance. With the rejection of 105 project evaluations and 67 research projects for "serious methodological weaknesses" (p. 39), the researchers found a significant positive effect (58%) or a nonsignificant effect (41%) on student performance, based on 3 project evaluations and 9 research studies. The authors advised that with the requirement of learning two languages, a nonsignificant finding demonstrates the positive advantages of bilingual education.

> The research demonstrates that bilingual education and bilingualism improves, or does not impede, oral language development, reading and writing abilities, mathematics and social studies achievement, cognitive functioning, and self-image. In addition, there is empirical evidence that bilingual education programs improve school attendance. (p. 8)

Possibly the most publicized and controversial study to date appeared on the national scene in three volumes during 1977-1978. Known as the "AIR Report," authors Danoff, Coles, McLaughlin, and Reynolds (1977a, 1977b, 1978), of the American Institutes for Research received a contract from the U.S. Office of Education, Office of

Planning, Budget, and Evaluation, to assess the status of federally-funded bilingual education programs. Even before the first volume was out, staff members Arias, Delgado, DeProcel, and Irzarry quit, claiming inadequate design and methodology (Willig, 1984, p. 3). The results from this evaluation of 38 Title VII projects in eight states indicted bilingual education, influenced legislation, and infuriated critics. Non-Title VII Hispanics outperformed Title VII Hispanics in English proficiency, while the reverse was found true for mathematics for the interim report, but corrected to equal for the final report. Two-thirds of Title VII classroom space was taken up with non-limited-English-speaking students who failed to exit when proficient. Finally, Title VII Hispanic students evidenced higher proficiency in Spanish than the non-Title VII control group. Critics such as Lopez and Cervantes (1978), Cardenas (1977), Gray (1977), O'Malley (1978), and Willig (1982) challenged the AIR Report's design and general findings. These findings, however, stimulated changes in the Educational Amendments of 1978, limiting implementation for any project to five years, and broadening the definition of English proficiency for LEP students, and setting an enrollment requirement of 75% or more LEP classified students.

Lyon, Dosher, McGranahan, and Williams (1978) looked at 116 final evaluation reports to assess whether they met minimum criteria for good program evaluation. The results indicated that evaluation reports frequently do not conform to simple standards. Of the 116 reports, 62% did not describe the program well enough so that objectives were clear; 90% did not address questions of validity and reliability of data sources; the data collection sources in 54% were not comprehensive enough to answer the evaluation questions; and 72% did not show congruence between information provided and conclusions.

Troike's (1978) monograph made a strong case for federal funding to develop an adequate research base for bilingual education while noting that, although evaluations should provide evidence for program results, "the vast majority of them (program evaluations) are worthless for this purpose" (p. 3). He then cited an earlier meta-evaluation of 150 final evaluation reports, only 7 of which met minimal criteria for acceptability. Troike then proceeded to discuss 12 unpublished evaluation studies and made the following pronouncement: "Despite the lack of research and the inadequacy of evaluation reports, enough evidence has now accumulated to make it possible to say with confidence that quality bilingual programs can meet the goal of providing equal educational opportunity for students from non-English-speaking backgrounds" (p. 3). Troike has been criticized by Baker (1984) for incorrect data, questionable interpretation of data, and citing of a project evaluation that no one else can find. "A good example of the misuse of research in the pursuit of advocacy of bilingual education is provided by Troike (1981)" (Baker, 1984, p. 17).

Dulay and Burt (1979a, 1979b) reviewed 38 research projects and 175 project evaluations but found that only 12 studies, 9 research studies, and 3 bilingual program evaluations met minimum research design standards. However, from those 12 studies, 66 findings were generated, with 58% showing positive effects as a result of bilingual education, 41% showing no effects, and only 1% showing negative effects. Bilingual education was considered successful by Dulay and Burt because it either improved or did not hinder academic achievement in school.

Martin (1981) completed a meta-evaluation of the final evaluations of 43 out of a total of 128 Title VII projects in the State of California. The results of this meta-evaluation study provided a description of evaluation design characteristics, data collection procedures, testing instruments in use, data analysis techniques, and general evaluation methodologies. Another author (Willig, 1984) used a specialized form of meta-evaluation, called meta-analysis, where statistics from the sample of evaluations were aggregated and reported in a metric called effect size. Willig conducted a reanalysis of the 23 studies analyzed by Baker and de Kanter (1981) under the auspices of the Office of Planning, Budget, and Evaluation, U.S. Department of Education. Baker and de Kanter used the traditional vote score method for aggregation of data while Willig calculated effect sizes. Baker and de Kanter's study, which received a good deal of media attention, concluded that the case for bilingual education was weak and that there was no justification for assuming that it is necessary to teach nonlanguage subjects in the students first language in order for the LEP student to make satisfactory progress. Baker and de Kanter's research was criticized by Littlejohn (1981), Rotberg (1983), Seidner (1982), Yates and Ortiz (1983), and Willig (1982) on methodological grounds and possible bias in selection of studies. Willig's results indicated that, contrary to Baker and de Kanter's results, there were small to moderate differences favoring bilingual students over comparison students for reading in English, language skills in English, mathematics in English, and total achievement in English, when the effects of the other variables were held constant. Willig attributed the differences in results to meta-analysis' greater power to detect true differences between groups due to the accumulation of estimated effect sizes over students and to make possible statistical controls for variables in the final analysis (Willig, 1984, p. 116).

Finally, the last study to be considered was also a meta-analytic synthesis of Title VII evaluations and research reports from the years 1977-1981, but reported in 1983 by James English, Office of Planning, Budget, and Evaluation, U.S. Department of Education. English had major difficulties locating documents, only finding approximately 60%. Of the reports available, only 12% of those prior to 1980 and 25% from 1980-1981 were deemed qualified to be included. English recommended that the Department of Education increase its internal support for project evaluation and provide technical assistance including a more formalized evaluation format.

Zappert and Cruz (1977) commented that

> too often, bilingual education programs which are doing an excellent job of motivating limited- and non-English-speaking students, involving parents, and succeeding in

> improving academic achievement of students are not adequately assessed because of a poor evaluation, research designs, . . . all factors which obscure the significant positive effect of bilingual programs. (p. iv)

Boruch, Cordray, and Pion (1981) agreed, "the evidence presented in the majority of evaluations is often insufficient for judging the effects of the projects or programs on children" (p. 17). Finally, Baker and Pelavin (1984) contributed the fitting summary to this review of bilingual evaluation studies by stating, "on one point all reviews agreed: the overall quality of bilingual evaluation research [emphasis added] is very poor" (p. 1).

Evaluation Research Versus Program Evaluation

In the previously stated remark by Baker and Pelavin (1984, p. 1), the underlining of the word research did not appear in the original quotation, but is added for emphasis. The federal government and many of the previously cited researchers have not made a distinction between evaluation research and program evaluation. Stufflebeam (1978) stated that "only chance renders an information system adequate when the decision maker does not specify the expected decision situations to be served by the system" (p. 123). Sanders (1981) continued, "information collected for a particular purpose at one level may have little or no utility at other levels of a national system of projects" (p. 3). Willig (1984) added,

> in short, the implications for research policy suggest the necessity of making a clear distinction between program research and the kinds of evaluations that are needed by school districts. . . . Districts should not

> be put in the position of attempting to answer a question that is truly impossible to answer in non-research settings, i.e., what are the effects of this program on participants compared to what the effects would be without such a program? (p. 115)

What we have seen is the post hoc expectation of the federal government and, many researchers in the field as well, that research data proving the effectiveness of bilingual programs should be available by collecting and synthesizing any quantity of LEA program evaluations originally prepared to meet the information needs of the local project. The previously mentioned reviews have illustrated that such a plan puts an undue expectancy on a system that is not adequately prepared to handle it, nor provided with the technical assistance to accomplish it.

Lincoln and Guba (1984) distinguished program evaluation from evaluation research by purpose, products, outcomes, and audiences. De George (1983) saw evaluation as

> a device intended to inform the manager and his/her staff to what degree they have accomplished what they have set out to do and how they can improve it. Because it is the vehicle whereby that same manager and staff account for their stewardship and the basis on which funding agencies and sponsors will decide whether to continue or terminate the program, clearly, evaluation is more than research. (p. 6)

Lipsey, Crosse, Dunkle, Pollard, and Stobart (1985), after wrestling with the question of program evaluation and the experimental paradigm, came to the following conclusion:

> Eventually, it may be wise to distinguish between program research which will draw on the power of the experimental paradigm for careful application of the causal links and

> theoretical propositions embodied in social programs, and program evaluation, which will use less restrictive approaches to provide prompt, useful information to policy makers and program administrators about the specific programs with which they are concerned. By its very nature, program research may not be amenable to routine application under circumstances where answers are needed quickly or where support for extensive programmatic research is not feasible. And, by its nature, program evaluation may not be capable of answering the ultimate causal question, "Does this social intervention produce the intended effects?" The challenge to the evaluation profession is to know the difference between these approaches and their domains of applicability. (p. 26)

Several solutions have been proposed to solve the dilemma of the need for research to judge the effectiveness of bilingual programs to report to congress and program evaluation to provide data to meet the information needs of the local bilingual education project. Willig (1984) proposed the establishment of federally-funded district research centers to conduct research in bilingual education, where students and teachers could be randomly assigned and where threats to validity and reliability could be controlled. These centers would operate independently of the LEA projects where program evaluations would seek to meet the information needs of the local decision makers. Millman (1981) proposed the distinction between a federal reporting system and an evaluation system. The reporting system would provide summative information about projects to address a specific federal need for information. The data would be comparable across projects and in a form for aggregation nationwide. The program evaluation system would be tailored to the idiosyncratic nature of the particular project. Sanders (1981) suggested that the experience gained from Title I program monitoring be transferred to the monitoring of bilingual accomplishments.

The preceding discussion is timely because the federal government is currently considering alternative procedures and materials for the evaluation of bilingual programs. In July, 1985, SRA Technologies, Inc. was awarded a contract to design and develop a new system for the evaluation of bilingual programs. Field testing is expected to begin in 1986-1987.

The question of research versus program evaluation is important to this meta-evaluation study. Although it is critical to the development of a new system to have an indepth understanding of the current status of LEA program evaluations, the major focus of this meta-evaluation was on program evaluation for the local decision makers.

CHAPTER III
METHODOLOGY

The purpose of this study was to conduct a meta-evaluation of Florida Title VII LEA bilingual education project evaluations, fiscal year 1984-85. This study examined evaluation model and design characteristics, data collection procedures, testing instruments, data analysis techniques, and general evaluation methodologies, as well as discrepancies between proposed evaluation plans and summative evaluation reports in order to answer the following four questions:

1. Which context and process variables are addressed in Title VII project summative evaluation reports within the State of Florida?

2. What are the characteristics of Florida Title VII evaluation models, designs, and reports in terms of information coverage, content, and procedures?

3. Do Florida Title VII evaluation reports meet the current accepted professional standards of good evaluation practices, as detailed by the Joint Committee on Standards for Educational Evaluation, _Standards for Evaluations of Educational Programs, Projects, and Materials_?

4. Are there major discrepancies between LEA proposals and summative evaluation reports? What is the nature of these discrepancies?

Using Stufflebeam's (1981a) terminology, summative meta-evaluation may be classified as a retroactive assessment of evaluation studies, their goals, designs, implementation, and results combined into a simple summary case study. Stufflebeam and Shinkfield (1985) further explained that "meta-evaluation invokes the accepted standards of the profession and assesses and tries to assure that they are met" (p. 35). The aim of meta-evaluation, according to Stufflebeam and Shinkfield, is "to assure quality evaluation services, to guard against or deal with malpractice or services not in the public interest, to provide direction for improvement of the profession, and to promote increased understanding of the evaluation enterprise" (p. 34).

Conceptualization of Meta-Evaluation

Daniel Stufflebeam (1974a, 1974b) was not only the first researcher to define the basic concepts of meta-evaluation, but even today his conceptualization remains the standard. Stufflebeam posited that the conceptualization of meta-evaluation must be consistent with an evaluator's conceptualization of evaluation. The following eight premises for the conceptualization of evaluation and meta-evaluation were outlined in Stufflebeam's (1974a) monograph, Meta-Evaluation:

1. Evaluation is the assessment of merit; thus, meta-evaluation means assessing the merit of evaluation efforts.
2. Evaluation serves decision making and accountability; formative meta-evaluation pro-actively provides information for decision making, while summative meta-evaluation retroactively provides information for accountability of past evaluation work.
3. Evaluations should assess goals, designs, implementation, and results; thus, meta-evaluation should assess the importance of evaluation objectives, the appropriateness of evaluation designs, the adequacy of implementation of the designs, the technical adequacy of data analysis, and the quality and importance of evaluation results.
4. Evaluation should provide descriptive and judgmental information and appropriate recommendations; thus, meta-evaluation should describe and judge evaluation work and should provide recommendations for improvement and utilization of evaluations.
5. Evaluation should serve all persons who are involved in and affected by the program being evaluated; hence, meta-evaluation should serve evaluators and all persons who are interested in the work.
6. Evaluation should be conducted by both insiders (generally formative evaluation for decision making), and outsiders (generally summative evaluation for accountability). Hence, evaluators should conduct formative meta-evaluation and they should obtain external judgments of the overall merit of their completed evaluation activities.
7. Evaluation involves the process of delineating the questions to be addressed, obtaining the needed information, and using the information in decision making and accountability. Hence, the meta-evaluator must delineate the specific meta-evaluation questions to be addressed; collect, organize, and analyze needed information; and apply the obtained information to the appropriate decision-making and accountability tasks.
8. Evaluation must be technically adequate, useful, and cost effective, and meta-evaluation must satisfy the same criteria. (Stufflebeam, 1974, pp. 70-71)

Sample

The population for this research project included all Title VII bilingual education projects in the state of Florida which received federal funds for the fiscal year 1984-85. According to records in the Bilingual Consultants Office, Florida Department of Education, Tallahassee, Florida, 14 new or continuing Title VII basic and demonstration grants were funded in the fiscal year 1984-85.

The sample consisted of Title VII projects which voluntarily contributed either or both application proposals and final evaluation reports.

Data Collection Procedures

The data set was confined to application proposals which included a proposed evaluation plan and final evaluation reports from the sample. Most Title VII proposals are kept on file in the Bilingual Consultants Office, Florida Department of Education. Those proposals that were not on file were mailed directly to the researcher by the project managers. Proposals are considered sensitive documents by districts because Title VII Bilingual Education Basic Grants are awarded on a competitive basis, discouraging intra-district collaboration on proposals. One district refused to submit a proposal, citing past negative experiences with proposal material being used by other districts without permission. Therefore, 11 out of 12 proposals were included in the data set.

The final evaluation reports, written by the project evaluator, were requested directly from the project managers, with 12 out of a total of 14 final evaluation reports received by the researcher. One district sent a formal letter requesting not to be included in the sample. Two districts, which employed the same external evaluator as the district which requested not to be included, had not submitted final evaluation reports as of May 8, 1986, for their 1984-85 grants award period. Both project managers agreed to supply the reports to the meta-evaluation if and when they did receive them. One arrived May 30, leaving the second of the two to be omitted from the data set. It must be noted that federal regulations require final evaluation reports to be submitted to OBEMLA in Washington, DC not later than 90 days after the termination of the grant, placing the deadline in December, 1985.

Instrumentation

No one instrument currently exists which synthesizes all of the criteria advocated in the multitude of models, checklists, standards, guidelines, pitfalls, and recommendations found in the literature for conducting meta-evaluations. Indeed, if all of the elements of the various aforementioned documents were synthesized into one meta-evaluation instrument it would be unwieldly and ineffective for any meta-evaluation. However, one set of instruments developed by Martin (1981) has the advantage of being specifically designed, field-tested, and employed in a meta-evaluation of 47 Title VII project final evaluations for the year 1979-80. The instruments represent a careful synthesis of the following 11 documents:

1. Key Evaluation Checklist (Scriven, 1974)
2. Illustrative List of Pitfalls in Evaluation Works (unknown)
3. An Administrator's Checklist for Reviewing Evaluation Plans (Stufflebeam, 1974a)
4. Meta-Evaluation Criteria (Stufflebeam, 1974a)
5. Objectives of the Meta-Evaluation (Stufflebeam, 1974a)

6. Checklist for Judging the Adequacy of an Evaluation Design (Sanders & Nafziger, 1976)

7. Table of Contents for a Final Evaluation Report (Stake, 1983)

8. Suggested Title VII Evaluation Design Specifications (Stake, 1983)

9. Suggested Title VII Evaluation Design Specifications (Office of Program Evaluation and Research and Bilingual-Bicultural Education Section, California State Department of Education, October, 1977)

10. Checklist for Planning and Managing the Evaluation of Title VII Projects (OPBE, U.S. Department of Education)

11. Checklist for Planning the Evaluation of Multi-funded Programs (Office of Program Evaluation and Research, California State Department of Education).

Since their development, no other researcher has employed Martin's (1981) instruments in a Title VII meta-evaluation. The present study used all four of Martin's instruments for initial data gathering purposes.

Instrument #1, the "CES Meta-Evaluation Checklist" (Martin, 1981), contains 153 items and provides for an indepth analysis of criteria related to evaluation model characteristics, information coverage, general and specific design characteristics, data collection techniques, testing schedules, sampling techniques, tests, analytic procedures, and evaluation utility.

Instrument #2, "Program Design Data Sheet" (Martin, 1981), contains 102 items pertaining to such program characteristics as size, number of schools, grade levels, and languages served, staffing, evaluation resources, program emphasis, program type, instructional approaches, student achievement, staff development, community involvement, curriculum evaluation, and materials in use.

Instrument #3, "Supplemental Data Sheet #1" (Martin, 1981) is modeled after Alkin, Kosecoff, Fitz-Gibbon, and Seligman's (1974) study rating overall quality of various data collection and analysis techniques in Title VII evaluation reports. This is a 51-item checklist rating items on a 1-through-3 scale. The items are related to the general organization and presentation of evaluation findings, overall proposal coverage, overall evaluation design coverage, assessment techniques, data analysis, and report characteristics pertaining to validity, objectivity, credibility, and utility.

Instrument #4, "Supplemental Data Sheet #2" (Martin, 1981), is reserved for extra narrative comments regarding model characteristics, design characteristics, and special notes on data analysis techniques. There is also space provided for comments on exemplary practices or areas in need of improvement.

The present researcher designed an instrument based on the 30 standards for evaluations of educational projects, developed by the Joint Committee on Standards for Educational Evaluation (1981). This instrument was used in judging the merit and worth of evaluation effort in the sample proposals and final reports.

Data Analysis

Meta-evaluation procedures as outlined by Stufflebeam (1974a) were used in the analysis of data. The retroactive meta-evaluation of goals, designs, implementation, and results of the 1984-85 Florida Title VII bilingual education project evaluations were accomplished through a careful application of the four previously described meta-evaluation instruments (Martin, 1981) and the instrument based on the standards developed by the Joint Committee on Standards for Educational Evaluation.

Data collection ranged from a simple check on a checklist type instrument indicating whether specific items were or were not addressed in each proposal and final report to copious notes describing and rating the quality of the primary evaluator's report in terms of such factors as comprehensiveness of coverage, timeliness, and technical adequacy.

Analysis of the meta-evaluation data employed descriptive statistics to characterize various aspects of the sample of evaluation designs and reports. Frequency counts, means, modes, and percentages were calculated where appropriate to identify the most common characteristics in the designs and reports. Averages were used to determine the extent to which various context and process variables were developed and objectives implemented. Discrepancy analysis was employed to determine if the percentage of discrepancy (%d) between the evaluation proposal and the final evaluation report was significant.

The results of the various analyses which make up the meta-evaluation were employed to answer the four main research questions. A listing of variables and characteristics were required to answer questions 1 and 2. In question 3, meta-evaluator judgment was required to determine if the sample of final evaluation reports did or did not meet each of the 30 standards of good evaluation practices. A combined sample score of 80% on all relevant variables or criteria was used as the cut-off point. A score of 80% or higher was required to assess that the standard was met. A description of the nature of the discrepancies between the application proposal evaluation plans and the final evaluation reports was required to answer question 4. The results of questions 1, 2, 3, and 4 are presented in Chapter IV.

CHAPTER IV
RESULTS

The purpose of this study was to conduct a meta-evaluation of Florida Title VII LEA bilingual project evaluations, fiscal year 1984-85. This chapter presents the meta-evaluation findings to the following four questions:

1. Which context and process variables are addressed in Title VII project summative evaluation reports within the State of Florida?

2. What are the characteristics of Florida Title VII evaluation models, designs, and reports in terms of information coverage, content, and procedures?

3. Do Florida Title VII evaluation reports meet the current accepted professional standards of good evaluation practices, as detailed by the Joint Committee on Standards for Educational Evaluation (1981), Standards for Evaluations of Educational Programs, Projects, and Materials?

4. Are there major discrepancies between LEA proposals and summative evaluation reports? What is the nature of these discrepancies?

Question 1: Context and Process Variables

Final evaluation reports should include sufficient descriptive information that they can stand on their own merit without reference to proposals or other documentation. Context variables are included in an evaluation to give the reader a clear picture of the setting in which the project functions and the unique features of the project being evaluated. Sufficient contextual variables should be addressed in a final evaluation report so that the reader may judge the effect of the contextual conditions on the project and under what similar conditions the findings may be applicable. Examples of context variables include geographic location, community characteristics such as ethnic composition and migration, a description of the district, and a description of how the project fits into the local school sites.

Program characteristics provide a picture of the unique features of the project such as the number of students receiving treatment; the number of sites: the ethnicity of students; program type, design, and instructional approaches; program emphasis; and staffing. Process variables should describe the implementation of the project goals, management of the project, timeline for activities, staff development, materials, progress records, and community involvement.

The first meta-evaluation question specifically asked which context and process variables were addressed in Title VII project summative evaluation reports within the State of Florida? Although 14 bilingual projects received Title VII funding in 1984-85, only 12 projects submitted final reports to the meta-evaluator. Data from all 12 final reports were analyzed and the results are presented in the order of context variables, program characteristics, and process variables.

Context Variables

Geographic location. Only two projects, representing 16.7% of the sample (N=12), discussed the location of the project in the narrative. A total of 10 reports (83.3%) listed the county name on the cover, but in five of these reports (41.7%) the cover was the only indication of the location of the project. An additional five reports mentioned the country name in the narrative, but two of these reports (16.7%) did not identify the name of the county on the cover.

Description of district. Three final reports (25%) included a brief description of the district. Five reports (41.7%) included a count of limited-English proficient (LEP) students for the whole district, while these five plus an additional three reports, making a total of eight (66.7%) discussed the project student population by such factors as ethnicity and grade, ethnicity and school, school and grade, exceptionality and school, ethnicity, school, and grade. Seven reports (58.3%) discussed socioeconomic factors of the student population such as parents'educational level, parents' occupations, and percent of students receiving free and reduced lunch.

Available resources. Six reports (50%) listed district contributions to the project, such as administration, staff, housing for the project, and internal evaluator. One district contracted with a university for services and those services were described.

Community characteristics. The most frequently mentioned community characteristic was ethnic composition, with 10 reports (83.3%) listing the following combinations: Hispanic in general, Puerto Rican, Haitian, Asian in general, Vietnamese, Thai, Laotian, Korean, Chinese, Greek, Tagalog, Hebrew, European in general, French, Rumanian, and Black. One report characterized a community as a long time Anglo community with recent immigration of Puerto Ricans to one area of the county. Migration was mentioned as a factor in five reports (41.7%), mainly stimulated by agricultural workers following the crops and in one instance a military base and a refugee resettlement center. One report (8.3%) calculated length of time in the U.S. and another report (8.3%) calculated number of days in school by nationality and migrant/non-migrant status. Four reports (33.3%) mentioned the socioeconomic factors connected with occupations and unemployment within the community.

Local school context. Not a great deal of information was provided on the Title VII project within the local school site. Nine reports (75%) did address site selection, how or why sites were selected to house the bilingual program. Seven reports (58.3%) provided the information that Title VII projects were housed in sites that already had bilingual programs established. Seven reports (58.3%) identified individual school sites by name, either in the body of the report or appendix, one report identified sites by number, and four reports (33.3%) did not identify sites at all.

Program Characteristics

Project type. Two projects (16.7%) were identified as demonstration projects, while the remaining 10 projects (83.3%) were either identified or assumed to be basic projects. Three projects (25%) were identified as new projects completing their first year and receiving their first evaluation. Six projects (50%) were identified as functioning under continuation grants, with five of the six (41.6%) identified as having completed their second year, however, there was confusion in one report which identified the current year as the second year at one point and as the fourth year when attainment of objectives was presented. The sixth report did not specify the year of the continuation grant. Three projects (25%) did not identify the length of time the project had been in operation, a serious oversight in that there are differing expectations for information coverage such as degree of implementation for first and final year evaluations.

Size of project. Seven of the 12 reports (58.3%) presented an unclear and confusing picture of either or both the number of students served or the number of project sites. With data available from only nine reports, a total of 2,380 students were served under Title VII grants in those nine sites, averaging 264 per site. The average number is misleading, however, since three of the projects were very large (844, 581, and 377) and the remainder were between 50 and 162 students served. Two projects did not provide service directly to students. One project provided training to parents of LEP students and the other project provided inservice training to 218 teachers, 68 aides, and 43 administrators.

Five projects (41%) provided services exclusively to elementary schools, two projects (16.7%) provided services to elementary and middle schools, three projects (25%) provided services to elementary, middle, and senior high schools, one project report did not identify number of sites or levels, and one project only provided training services to staff. Of the 10 project reports which recorded the number of sites servod, it was reported that 51 schools housed Title VII project services. Nine of the 10 project reports recorded the number of sites at each level, with 25 sites (49%) allocated to elementary schools, 12 sites (23.5%) allocated to middle, 7 sites (13.7%) allocated to high schools, and 7 sites (13.7%) unidentified. One of the 7 high school sites was in a private religious school setting.

Languages. LEP Lau category A and B students whose dominant language was Spanish received treatment in 75% of the projects, with four projects (33.3%) exclusively serving Hispanic students, three projects (25%) serving a majority of Hispanic and a minority of Haitian-Creole students, one project (8.3%) serving "Spanish, black, and others," and one project (8.3%) serving multiple languages including Spanish, Haitian-Creole, and Vietnamese. The remaining 25% consisted of one Greek/Anglo project, one project serving mostly Asian languages, and one project which did not identify languages.

Student characteristics. LEP students, whether or not they have previously attended school in their home country, are usually served in the same grade grouping as students with whom they are mainstreamed for art, music, physical education, and increasing amounts of content instruction as their comprehension of English increases. For this reason the majority of reports distinguished students by grade level, but one report (8.3%) of a project that specialized in providing services to exceptional students did identify students by age groupings. It is not uncommon, however, for a LEP student to spend more than one year at a certain grade level or more than one year in the bilingual project. Only one report (8.3%), however, presented student achievement data by number of years in the project. Three reports mentioned transiency as a factor but only one project (8.3%) presented the average number of days of attendance by language group and occupation. Finally, only one report addressed rate of retention, falsely labeling it as drop-out rate, with no reports addressing the drop-out rate.

Program type. Seven projects (58.3%) were described in the final evaluation reports as transitional bilingual projects, one project (8.3%) was described but not labeled as a maintenance program, one (8.3%) was referred to as an intensive English program, one project (8.3%) was not labeled, and the distinction did not apply to two projects (16.7%).

There is no uniformly applied and accepted definition of what constitutes a bilingual program either at the federal level or in the literature. Since there is no state legislation in Florida governing or defining bilingual education, there are as many variations in practice as there are programs in this state, which contributes to the difficulty of identifying and classifying a program using nationally recognized labels such as full-bilingual, partial-bilingual, or intensive English. In terms of the instructional approaches described in the final evaluations, five projects (41.7%) might be labeled as partial bilingual, while three projects (25%) might be labeled as full-bilingual programs, with one (8.3%) being identified as intensive English. However, if the label applies to the language of instruction of content subject matter, the three full-bilingual and two of the partial-bilingual programs (41.7%) presented content material in the student's first language while providing English as a second language (ESL) instruction. The three additional partial-bilingual programs (25%) provided content instruction in English and the first language as necessary. Finally, if the label applies to the classroom structure in which the program is delivered, the three full-bilingual programs were in self-contained bilingual classrooms except for art, music, and physical education. Two of these bilingual classrooms were staffed with bilingual teachers and one bilingual classroom, in what was described but not labeled as a maintenance program, was staffed with bilingual aides and a monolingual English teacher who alternated content instruction in first language and ESL with second language instruction to native English speakers and content in English to LEP students. The partial-bilingual programs ranged from half-day instruction in the elementary grades in self-contained bilingual classrooms (8.3%) with instruction in English language arts the remainder of the day to assignment in a mainstreamed classroom with a full-time bilingual aide (8.3%) with pull-out ESL to assignment in a mainstreamed classroom with an itinerant bilingual aide (25%) as needed with pull-out ESL. Middle and senior high school programs were described as providing a much more standard program of from one to two hours of ESL with tutorial help in content classes as necessary from bilingual aides.

The final evaluations for five projects (41.7%) mentioned the use of individualized instruction but specific bilingual or ESL teaching methodologies were not addressed in any evaluation report. Five projects (41.7%) did address the question of amount of ESL instruction time.

Entry and exit criteria. How a student qualifies to receive treatment in the Title VII project and what proof and procedures are required for moving a student from the project to mainstream English classes was addressed in seven final evaluation reports (58.3%). Three of the seven project reports (25%) simply noted that such criteria were in existence but four reports (33.3%) actually described specific criteria.

Program emphasis. Two projects existed to provide training to staff in bilingual centers and to parents of LEP students. The remaining 10 projects emphasized the following goals in descending order of frequency of selection: instruction (10), staff development (5), materials development (4), parent/community involvement (4), multicultural awareness (4), self-concept (2), parent training (2), management goals (2), guidance (1), identification (1), and computers (1).

Staffing. Six project reports (50%) made some reference to staff qualifications. Of the 12 project reports, 10 (83.3%) made some attempt to identify staff positions but the data were unclear as to the total number of staff positions, those paid by Title VII, and those paid by the district. It was also unclear as to which staff members were bilingual and which were monolingual English. Most reports did not address the qualifications or amount of inservice received by mainstream classroom teachers with and without bilingual aides. The qualifications of instructional bilingual aides was not addressed. Most evaluation reports did not address the question of whether the project director was a trained bilingual educator or a district administrator.

Process Variables

Implementation evaluation. Certifying whether a project has been implemented as proposed and documenting when each aspect of the project was set in place as fully functioning enables the evaluator to determine whether or not the project is evaluable. Evaluation may not be valid or reliable if a project has incurred major delays in obtaining and training staff, providing materials, assigning students to all sites, standardizing curriculum, teaching methodologies, test administration, or other similar problems. Determining if a project has been sufficiently implemented to warrant a full evaluation should be of prime consideration for new projects. The evaluation of continuing projects should, as a matter of course, present evidence through observation, interviews, questionnaires, or appropriate documentation as to how the project is actually functioning in all sites. Evaluation based on the assumption that a project is functioning according to proposal specifications may be no more reliable than a public relations document.

Two final evaluation reports (16.7%) out of the sample of 12 reports completed a thorough implementation evaluation as an integral part of the final report. One of the two projects was identified as being a first-year project while the other report did not specify that information but consisted entirely of a process evaluation conducted by an internal evaluator who did not address product questions of student achievement. Four other projects out of the total of 14 awarded in Florida in 1984-85 were in their first grant year. One grant administrator did not submit a final report to the meta-evaluator, one report was submitted but did not identify the project as being in its first year and two other submitted reports did identify the project as being a first year grant but did not provide an implementation assessment.

Although an activity timeline is one step in the right direction, it alone can not be considered an implementation evaluation. Four reports (33.3%), one of which identified the project as being in its first year, did present in chart form the project's main objectives, proposed activities, and a timeline consisting of dates of planned and actual accomplishment. This left six project reports (50%) which did not address the question of implementation in any form.

Staff development. The inservice training of Title VII bilingual teachers, bilingual instructional aides, and mainstream classroom teachers was addressed in nine reports (75%) with one project (8.3%) designed specifically to install an instructional management system and to provide inservice training to locally funded bilingual program staff and regular classroom teachers with LEP students. Seven (58.3%) of the nine projects which addressed staff development tabulated the number of inservice presentations,which ranged from 8 to 58, while five projects (41.7%) provided a listing of program types and topics. Five projects (41.7%) presented the participants' assessment of the value of the staff development activities, while two projects (16.7%) addressed the staff's application of acquired knowledge to the classroom situation. Two projects (16.7%) also reported providing training to parents and the community. Four projects (33.3%) provided college and university courses for bilingual staff and in one project provided university courses to parents with the aim of preparing a well-trained pool of applicants for future aide positions.

Materials. Acquisition of instructional materials was addressed by 10 projects (83.3%) but there was little agreement on topics covered. One project which did an implementation evaluation documented that materials arrived in good time to be used throughout the project year and listed the films available in both Spanish and English that had been added to the district film catalog during the year. Another report discussed the development and evaluation of computer software. Seven reports (58.3%) addressed materials development, adaptation, and translation, while three references (25%) were made to ordering published materials. The reports were not specific with regard to the language of the materials discussed although three projects (25%) did mention materials in Spanish and Greek and several other districts mentioned using regular district-adopted English texts.

Curriculum. The development, revision, or adaptation of curriculum was discussed in five reports (41.7%). One project requested funds for a three summer sequence of curriculum development and evaluation by a committee of teachers who received a stipend paid out of Title VII funds. Another project presented evidence of the revision of the ESL curriculum for use in computer assisted instruction.

Record keeping. Various systems for accumulating and storing accurate project records were discussed by eight project reports (66.7%). Instruments such as various types of pupil progression plans and continuums were identified in all eight reports. One district received Title VII funds to install a management information service for keeping student data which provided an individual student "map" of language skills mastered. Funds also covered the inservice training for bilingual staff and mainstream classroom teachers to ensure appropriate installation and use of the system. Other districts explored various ways to place student records on computer.

Parent participation/community involvement. In order to make application for Title VII funding a district must present proof of at least one parent meeting advertised in a local newspaper to inform the parents and the community at large of the proposed application and seek support for the project. Once the grant is awarded, a parent advisory council (PAC) is required. Provisions are made so that non-English-speaking parents will be able to understand and contribute to the meetings. Since these are Title VII requirements, the expectancy was that some aspect of parent participation would be mentioned in all of the final evaluation reports, however, various aspects of this topic were addressed by only seven project reports (58.3%). Five project reports (41.7%) mentioned PAC meetings with several reports appending bilingual invitations to meetings and at least one attendance "sign-in" sheet. Parent education was the main focus of one project and three others (totaling 33.3%) discussed the provision of some form of parent education although one of the three was not able to provide the parent education services it had proposed. The parent education project included as a portion of its curriculum the preparation of non-English-speaking parents to become involved in community and school activities and to provide volunteer services to the school. Another report mentioned a needs assessment and a third report indicated parent-community members who had taken college courses through project sponsorship as previously mentioned under staff development.

Summary of Results for Question 1

Question 1 asked, "which context and process variables are addressed in Title VII project summative evaluation reports within the State of Florida?" Implied in this question is the notion of adequate representation, so that the question might be rephrased as "which context and process variables are most frequently represented in the Title VII project summative evaluation reports within the State of Florida?" The ethnic composition of the community was the most frequently mentioned context variable, followed in descending order by site selection in the local school context, project student population in the description of the district, district contributions in the available resources, and geographic location.

The languages represented in the student population were the most frequently reported program characteristic, followed in descending order by identification of program emphasis, project type, entry and exit criteria, staffing, program type, size of program, and student characteristics. Materials were the most frequently mentioned process variable, followed in descending order by staff development, record keeping, curriculum,and implementation evaluation.

Question 2: Evaluation Models, Designs, and Reports

Although some dictionaries may attribute synonym status to "model" and "design," they are considered as two quite separate entities in the specialized field of evaluation. It would not be a gross exaggeration to say that there are as many evaluation models as there are theorists in the field. The label model identifies a philosophically inspired conceptualization of evaluation and the methodologies, procedures, tasks, and roles which characterize that conceptualization. Baker (1980), in describing the formative evaluation model, said it formed "a basis against which to assess the quality of different kinds of instructional sequences" (p. 23). Question 2 attempted to discern whether the procedures used in evaluating Title VII bilingual programs in Florida were motivated by identifiable model conceptualizations or eclectic or random actions.

An evaluation design identifies the "conditions and schedule under which the measures are taken or the data collected" (Madaus, Scriven, & Stufflebeam, 1983, p. 122). Seldom are bilingual evaluators, working in the context of local school districts, able to impose the experimental design's requisite of randomization to sample and control group selection. This second question attempted to identify and describe the actual conditions under which evaluation data were collected.

Question 2 asked "what are the characteristics of Florida Title VII evaluation models, designs, and reports in terms of information coverage, content, and procedures?" The evaluation reports themselves were examined for such information as authorship, information coverage, procedures, and form and style of presentation. Data from the 12 final evaluation reports which were submitted to the meta-evaluation were analyzed and the results are presented in the order of evaluation reports, evaluation models, and evaluation designs.

Evaluation Reports

Evaluator. Three final evaluation reports (25%) did not identify the name of the evaluator on the cover, in the introduction, or in the body of the report. One of these three reports did make reference to the hiring of an external evaluator and identified her as "she," while the name of another evaluator did appear in a meeting agenda placed in the appendix. From the nine reports (75%) which did identify the evaluator, the following information was available: five reports (41.7%) were completed by Ph.D.s, three reports (25%) identified the evaluator's company name and address, one report (8.3%) was completed by a team of two Ph.D.s, two (16.7%) of the nine Title VII districts which identified the name of their evaluator hired the same external evaluator, and, finally, one named evaluator was also identified as the evaluator for the two districts which did not submit final evaluations to the meta-evaluation. Three Title VII projects (25%) out of the total sample of 12 used internal evaluators, 8 projects (66.7%) hired external evaluators, and no information appeared in 1 report (8.3%). Only two reports (16.7%) indicated the point of entry of the evaluator to the project, March and April, well after the projects had been placed in operation. The final evaluation report submission date was provided by seven reports (58.3%).

Final evaluation reports. An informative introduction to the final evaluation report giving such information as number of students, staff, curriculum, type of treatment, etc. appeared in eight reports (66.7%) while some information such as purpose and curriculum was provided in one introduction (8.3%). Two final reports were prefaced, one by an abstract, and one by a summary of results, but neither provided sufficient contextual and descriptive information. One report (8.3%) opened directly with the presentation of results.

Five reports (41.7%) clearly stated the purpose of the evaluation, one (8.3%) alluded to a purpose, while six reports (50%) did not address purpose. The objectives of the evaluation were specified by two of the reports which formalized a purpose statement.

Information coverage was fairly comprehensive in 10 (83.3%) of the 12 final evaluation reports with most reports presenting data on a broad range of program objectives and activities. One evaluation report (8.3%) described five different instructional components provided according to LEP student characteristics, but the majority of evaluation information presented pertained to only one of the five instructional components. Another evaluation report (8.3%) addressed only process variables. In response to an inquiry regarding a copy of the product evaluation, the district volunteered that no further evaluation had been done for the year 1984-85. The following topics covered in Title VII evaluations are presented in descending order of frequency: cognitive achievement--prespecified objectives, staff development, parent participation/community involvement, activities in relation to goals and objectives, affective prespecified goals, materials development, management, staff performance and attitudes, and curriculum development.

Seven of the 12 final evaluation reports (58.3%) closed with evaluator recommendations, while 5 reports (41.7%) provided no recommendations at all. Evaluation audiences were not identified in the majority of final evaluation reports with only two reports providing an abstract or separate summary of results. It was mentioned in two reports (16.7%) that the parent advisory council would review the final evaluation report.

Evaluation Models

An objective-based evaluation model was evident from 7 of the 12 final reports (58.3%); a goal-based model was evident in 1 report (8.3%). Results were presented in these reports following the statement of the objective or the goal. An additional two reports (16.7%) presented a statement of the evaluation question followed by the results. No report presented both the objective or goal and the evaluation question. Specific evaluation questions were not stated in 10 reports (83.3%). The process-product model was in evidence in five reports (41.7%). The formative-summative model was mentioned in two reports (16.7%), while seven reports (58.3%) mentioned final or summative only. Implementation evaluation was mentioned in three reports (25%), one objective-based/final, one question-based/process-product, and the goal-based/process-product report.

Evaluation Designs

Control groups. Randomization, the requisite of a true experimental design, was not accomplished in any of the 12 projects, although it was addressed in two reports. A control group was mentioned in one report as having been selected randomly from those LEPs whose parents did not accept training, while at the point in the report where the results were presented the control group was described as matched in characteristics to the target group whose parents did accept parent training. The evaluator treated this as one of Campbell and Stanley's (1963) true experimental designs--a pretest-posttest control group design. The evaluator's treatment does not qualify because there was no randomization; however, it appears to qualify as one of Campbell and Stanley's (1963) quasi-experimental designs--the nonequivalent control group design. Had the evaluator used the non-equivalent control group design, he/she would have had to deal with the problems of self-selection as a factor in the control and target group formation and regression to the mean, since both groups were selected originally for lack of progress and difficulty in passing the Statewide Student Assessment Test (SSAT). A second report mentioned that the school sites to be visited were selected at random but in a letter to the district principlas included in an appendix to the report it was stated that schools were selected from those not visited the previous year.

Comparison groups. The most frequently used comparison was the norm referenced design. Eight projects (66.7%) relied either in total or partially on the norm referenced design. The national norms supplied by test publishers was the rule, but two projects (16.7%) reported the use of locally established norms. One project reported a comparison between LEP target and LEP non-target students. Unfortunately for the design, all students in Lau categories A through E were identified as limited English proficient (LEP), a serious mistake, since by Lau definitions only categories A and B are limited English proficient.

Another serious error for the design occurred because all LEP students, target and non-target, received treatment although target students received "primary attention." A longitudinal comparison was discussed in three reports but employed by only one project which compared 1983-84 project scores with 1984-85 scores on the Comprehensive Test of Basic Skills. Unfortunately, the evaluator questioned the reliability of the 1983-84 scores, positing too much assistance from the aides. A third project compared the scores of those students who were in their second year of project instruction with those students who were in their first year of project instruction for each grade level. Comparison of student progress with a predetermined mastery level was employed by nine projects (75%) using project developed and published criterion-referenced tests.

Data collection: Tests. Although an evaluator may collect many types of data, both quantitative and qualitative, from numerous sources, the most frequently relied upon data were student achievement test results collected on a pretest-posttest basis from target students using a norm referenced model. Since the State of Florida has no legislation establishing or regulating bilingual education and since evaluators are usually not hired until a project has been funded and in operation, it is quite frequently the district personnel who write Title VII proposals who make critical decisions regarding test selection, data collection, and data analysis. Most reports (eight or 66.7%) did not include a rationale for instrument selection or include a description of the instruments used.

There does not appear to be much agreement in the sample of 12 reports regarding achievement test instruments used. Three projects (25%), all resident in the same school district, administered the California Achievement Test, two final evaluation reports (16.7%) noted the use of the Stanford Achievement Test, two (16.7%) reported results from the Comprehensive Test of Basic Skills, and one project each employed the following tests: Brigance Diagnostic Inventory of Basic Skills, IDEA Proficiency Test, Language Assessment Management System Tests, Science Research Associates Achievement Series, Stanford Diagnostic Reading Survey, Stanford Early Achievement Test, and the Wide Range Achievement Test. There is agreement on one test, the Statewide Student Achievement Test (SSAT). All Florida students who have been in English-speaking schools for two years and who are in the 3rd, 5th, 8th, and 11th grades are required to pass the Statewide Student Achievement Test for entrance to the next grade.

Although there is wide variation in instruments used, the procedures for screening students for entrance into Title VII bilingual projects are fairly standard. Students and parents are screened on enrollment regarding the language of the home and the languages of the student. When a language other than English is listed, the student usually undergoes an individual oral interview followed by a test of language dominance and quite frequently a test of language proficiency for placement or individualization of instruction. The following tests were listed for these purposes:

Crane Oral Dominance Test

Bilingual Syntax Measure I and II

Broward County Oral Assessment

Dade County Aural Comprehension

Dade County Oral Comprehension

Dade County Comprehensive Test

Dade County Elementary Language Development

Dade County Test of Language Development (Receptive)

Dade County Secondary Placement Test, ESL 7-12

Language Assessment Scales

Language Assessment Battery

Language Assessment Management System Diagnostic Tests

Language Assessment Management System Placement Indicators

Brigance Diagnostic Inventory of Basic Skills

Woodcock Language Proficiency Battery

IDEA Proficiency Test

Most projects used district-developed pupil progression plans, continuums, or published language management programs to record student accomplishment of specific skills. The Broward County Pupil Progression Plan was used by three projects and each of the following were used by at least one project: Highlands County Minimal Skills Screening Instrument, Language Assessment Management System, and IDEA Oral Language Management Program.

Two projects which tested for affective growth in self-concept reported using the Stenner and Katzenmeyer Self Observation scales

(Intermediate and Senior) and the Dade County Hispanic/Anglo American Cultural Attitude Inventory Test. Two districts used the Bilingual Vocational Oral Proficiency Test with older students who received vocational counseling. Bilingual exceptional students were tested with the following instruments: Leitner International Performance, Stanford Binet Spanish/English, Bender Gestalt Spanish/English, and System of Multicultural Pluralistic Assessment.

Five projects (41.7%) developed their own criterion referenced tests (CRTs) to meet special project needs. They were the following:

Student Academic/Behavior Assessment Inventory

State Assessment Tutoring Skills

Parenting Skills

School/Community/Cultural Relations

Bilingual Education

Computer Literacy

Criterion Reference Test of Language Proficiency (proposed to be validated by the Dade County Aural Comprehension test but validation not accomplished)

District produced Math Criterion Referenced Test

Transitional Bilingual Listening Skills

Transitional Bilingual Speaking Skills

County Math System Program--Greek/English

Greek Proficiency Tests

County Minimum Communication Skills Test

County Readiness Assessment Inventory, K

County Mathematics Skills Tests

Most tests were administered on a pretest-posttest basis, although four projects (33.3%) reported at least one posttest only test administration. Once passed the "short" fall-spring first year, most final evaluation reports which addressed testing schedules showed a marked preference for spring to spring testing with only one project (8.3%) reporting fall to spring testing.

Data collection: Questionnaires. Although tests accounted for the bulk of data results presented in the sample of 12 Title VII evaluation reports, 75% (9) of the final reports presented results from at least one questionnaire. Eight of the nine reports (66.7%) presented results of questionnaires to staff, both bilingual and mainstream, while only three reports (25%) each presented results from students or administrators and only two reports (16.7%) presented results from questionnaires to parents.

Data collection: Documentation. A multitude of information about Title VII projects can be gleaned from the various inclusions to the appendices of the final evaluation reports. Most of the material was included to support claims of meetings held, parent involvement, and such things as staff workshop titles and dates and newspaper clippings publicizing student activities. Unfortunately, the project evaluators in general did not make reference in their comments to the interesting information available at their fingertips nor to other appropriate available documentation that would have described the program as it actually functioned. Three reports (25%) did present data on student attendance patterns, promotion rate, and student retention rate which was wrongly described on the following page as the drop-out rate. One evaluation of a first year grant recipient, which included an implementation evaluation as part of the final report, made excellent use of documents. The evaluator examined team teacher-aide lesson coordination sheets and individual aide's lesson plan books for evidence along with aide and team teacher questionnaires to support the project's claim that there was coordination of teaching effort in content classes between the Title VII project and the mainstream classroom teachers. Other documents mentioned in that particular evaluation report were travel vouchers, the district's audio-visual catalog, purchase orders, and invoices for consultants and for bilingual psychoeducational evaluations. No reports looked at suspensions, disciplinary records, or drop-out rate.

Data collection: Interviews and observation. Four reports (33.3%) presented data from face-to-face interviews, with one mentioning phone interviewing as well. Administrators were the subjects of face-to-face interviews in all four evaluations while staff in general provided interview data in two reports (16.7%). No evaluation reported interviewing students, parents, or community members. Direct observation of the actual functioning of the Title VII project in its various classroom sites was reported in only two projects (16.7%) with evaluators not reporting personal observation data, first hand knowledge of the project they were evaluating, in 83.3% of the sample reports.

Data analysis. None of the final evaluations reported sophisticated statistical analyses. The majority of pretest-posttest data were analyzed for central tendency and significance. Seven projects (58.3%) reported pretest-posttest means and six (50%) tested the significance of the difference between means using a form of t-test. One of the seven projects used an analysis which calculated the median observed and expected scores per grade level on their district's pupil progression plan, testing significance with chi square. Four projects (33.3%) reported gains by percentage of students increasing posttest score over pretest score and labeling the column "percent increase." One of the projects added an additional column labeling it the percent of significant increase when the t-test was significant. Most reports did not present data on variability. One project presented mean score, standard deviation, and range per grade level for a project-developed criterion referenced language test. A second project presented separate analyses for 1983-84 and 1984-85, identifying posttest grade equivalent mean scores and standard deviations by grade level and subtest for program, non-program, and former program students on the Comprehensive Test of Basic Skills (CTBS). It was reported that significance was screened visually because of small sample size. The last of the three projects (25%). which addressed variability presented scale score means, standard deviations, and t-test results by project site.

Only one project (8.3%) presented correlational data. CTBS median scores were correlated with median scores from the Dade

County Aural Comprehension Test using Spearman's (rho) Rank Order Correlation. Finally, six projects (50%) calculated data using percentages. Of the six projects, one reported percentages, mean ratings, and range, and three projects (25%) used percentages exclusively as their only tool for statistical analysis.

Summary of Results for Question 2

Question 2 asked, "what are the characteristics of Florida Title VII evaluation models, designs, and reports, in terms of information coverage, content, and procedures?" The majority of reports organized the presentation of evaluation results by statement of project objectives or by instruments. When a model could be identified, the process-product model was used most frequently, followed by the summative-formative or just summative model.

The most frequently used design was the comparison group norm referenced design, followed by comparison to a criterion reference. Since no project successfully employed randomization, no true experimental designs were used, although one project design qualified as a quasi-experimental non-equivalent control group design. Although there was wide variation in the testing instruments that were employed, most instruments, normed and criterion referenced, appeared to have poor validity and reliability for the LEP student population and curriculum of the Title VII bilingual projects. Questionnaires were the most frequently reported source of non-test data, with notably infrequent use of documentation, interviews, and observation data reported. Data analysis did not employ sophisticated statistical analyses. Means and t-tests for significance were reported most frequently, followed by percentages.

Authorship of the final evaluation reports was not consistently indicated, and the evaluators did not provide information to establish their credibility. The reports were uneven in quality, evidencing weaknesses in description of context, program characteristics, and process variables, statement of evaluation questions, employment of a consistent model, employment of adequate data collection and analysis methods, and presentation of insightful evaluator recommendations.

Question 3: Professional Standards for Evaluation

Meta-evaluation, according to Stufflebeam (1974a), should address the merit of evaluation efforts to provide information for accountability of past evaluation work. Meta-evaluation should assess the importance of evaluation objectives, the appropriateness of evaluation designs, the adequacy of data analysis, and the quality and importance of evaluation results. Of major concern to the meta-evaluator is the selection of appropriate criteria for judging the meta-evaluation.

The Joint Committee on Standards for Educational Evaluation, headed by Daniel Stufflebeam and sponsored by 12 professional organizations, published 30 standards for good evaluation practices (Joint Committee on Standards, 1981). These 30 standards taken together form the basis for a working philosophy of evaluation, defining principles that should guide and govern evaluation and meta-evaluation efforts. Stufflebeam (1981b) recommended, "now that the field has articulated 30 standards of good practice, it will be important to ascertain the quality of evaluations in relation to each of the standards" (p. 43). (A listing of the 30 standards can be found in Appendix A.)

The explicit intent of this third question is to ascertain the quality of 1984-85 Florida Title VII final evaluation reports in relation to each professional standard. Question 3 asked, "do Florida Title VII final evaluation reports meet the current accepted professional standards of good evaluation practices, as detailed by the Joint Committee Standards for Educational Evaluation (1981), *Standards for Evaluations of Educational Programs, Projects, and Materials*?" This analysis follows the published order and grouping of the 30 standards: Utility Standards, A1-A8; Feasibility Standards, B1-B3; Propriety Standards, C1-C8: and Accuracy Standards, D1-D11.

Utility Standards

The following eight standards were designed to meet the practical information needs of given audiences.

A1--Audience identification. This standard was designed to ensure that the various individuals and groups who are involved in or affected by the evaluation are identified so their needs will be represented in the evaluation and so they will receive feedback from the evaluation results. The sample of 12 evaluation reports did not meet this standard. As was mentioned in question 2, only two reports (16.7%) identified any intended audiences and neither of these two reports provided a complete listing of those who should have been represented in the evaluation or who should have received feedback from the evaluation report. In the remaining 10 reports the question of audience was not addressed.

A2--Evaluator credibility. It is the responsibility of the evaluator to provide the client and the reader with a statement of his or her qualifications. Such documentation should establish integrity and competence in terms of training, technical skills, substantive knowledge, and experience which qualify him or her to perform the evaluation. The standard of establishing evaluator credibility in the final report was not met. As was reported in question 2, three reports (25%) did not identify the name of the evaluator, and nine reports (75%) did not identify a title, company name, or address. None of the reports provided a statement regarding the qualifications of the evaluator beyond the notation of a Ph.D. after the evaluator's name in five of the reports. One evaluator did, however, provide a statement certifying professional and ethical treatment of data and interpretation of results. Standard A2 was not met by the sample of 12 final evaluation reports.

A3--Information scope and selection. This standard addresses the comprehensiveness and selectivity of the data collected to answer the evaluation questions and meet the information needs of the various identified audiences. In order to meet this standard, unless a goal free evaluation model is specified, it is a requisite that the information needs of the various audiences be identified and that specific evaluation questions be designed in order to avoid a purposeless random collection of information and test results no matter how broad the spectrum or appropriate the data are that are collected. Given this premise, it must be judged that the sample of 12 final evaluations did not meet the standard, although information coverage was judged as fairly comprehensive in question 2. Data reported in question 2 confirmed that audiences and their information needs were not identified, the purpose for the evaluation was not stated in 50% of the reports, and specific evaluation questions were not formulated in 83.3% of the reports, although 66.7% of the reports substituted project objectives for evaluation questions.

A4--Valuational interpretation. This standard involved the value interpretations made on the significance of the information collected in the evaluation, identifying who made these value interpretations and on what basis the interpretations were made. In the perspective of a Title VII project, for example, it would be a serious error to assume that the objectives of the bilingual teachers and staff were the only values that needed to be taken into account, without considering the value judgments of parents, administrators, and mainstream teachers who either have LEP students a portion of each day, or who receive students when they exit the program. Unless a decision to the contrary is made with the client, it is usually the evaluator who takes all of the competing values into account, offers alternative bases for interpreting findings, and pronounces a judgment of merit or worth. One aspect of this standard was met since the reader is able to discern that it was the evaluator who made the value judgments in all 12 of the final reports. However, only seven (58.3%) of the evaluators provided recommendations based on the value interpretations they placed on the data collected, only one (8.3%) of the evaluators suggested alternative bases for interpreting findings, and only two (16.7%) of the evaluators issued specific recommendations to the projects they evaluated. In one instance the evaluator persisted in making a value judgment when attrition reduced the treatment group from 41 subjects completing the first cycle to 10 of those same subjects completing cycle two. Data were collected on the total group and a valuation interpretation was attempted even though 75% of the subjects had not received the full treatment in cycle two. Finally, at least 50% of the final reports were based on the value judgments of the project staff, without consideration of the values of other relevant audiences such as the administration, mainstream teachers, or parents. For these reasons it must be judged that the standard was not met.

A5--Report clarity. The final evaluation report should describe the project, the context in which it functioned, its purposes, procedures, findings, and recommendations. The report should be written in a concise narrative style, organized consistently, with appropriately captioned and understandable tables and charts.
The sample of reports taken as a whole contained many recommended features, such as a clear, professional narrative style, professional looking cover and/or paper stock, computer generated charts, graphs,

and tables, good contextual information, thorough description of program, documentation of implementation, statement of evaluation purpose, and evaluation questions. Unfortunately, the individual reports were all lacking in some critical elements. For example, the report with the most professional looking charts, graphs, and tables also had the most questionable data. A thorough presentation of contextual variables and program description was discussed under question 1. Four reports (25%) exhibited multiple problems, from careless errors to questionably inflated numbers of subjects actually receiving treatment to the basing of a case for success on a test reported as significant at the .05 level in the summary when the narrative and table presenting the data six pages earlier clearly showed that the difference was "not significant." Standard A5, Report Clarity, was not met by the sample of Title VII final evaluation reports.

A6--Report dissemination. A negotiated dissemination plan specifying editorial control, audiences for the evaluation results, and formats for presentation of results, and rights to release results was recommended by the Joint Committee. It was also the position of the Joint Committee that special effort be made by the client and evaluator to inform all "right-to-know" audiences and that, in the cases of withholding of evaluation results or misrepresentation of evaluation results, the evaluator bears the responsibility for informing the audiences, since the evaluator is ultimately responsible not only to the client but to the right-to-know audiences. Right-to-know audiences are defined by the Joint Committee as those who are entitled to be informed about the results of the evaluation for the following reasons: they are the client; they commissioned the evaluation; they bear legal responsibility for the project; they funded the project through taxes or gifts of time or money; they supplied data; they are stakeholders such as staff, parents, students, and representatives of the mass media.

Although the State of Florida has enacted "government in the sunshine laws" which validate by statute the public's right-to-know, this standard was frequently both passively and actively ignored in Florida Title VII project evaluations. The collection of final evaluation reports for the meta-evaluation is a perfect example of both passive and active ignoring of the standard and the law. A sample of 12 out of 14 final evaluation reports were ultimately collected over a five-month period. One project reported that they had not received a copy of the final report by May 8, 1986, five months after the report was due in the U.S. Department of Education, Office of Bilingual Education and Minority Language Affairs (OBEMLA). The 14th project finally requested in writing not to be included in the sample.

No final evaluation reports identified right-to-know audiences. Two projects (16.7%) mentioned sending the reports to OBEMLA and one project discussed submitting information from this evaluation and the results from the evaluations of similar projects to the

Joint Dissemination Review Panel. Standard A6, Report Dissemination, was not met by the population of 14 Title VII final evaluation reports.

A7--Report timeliness. Timeliness is usually defined as providing delivery of the final report so that the information can be used in planning and decision making. The Title VII fiscal year for 1984-85 extended from October 1, 1984 through September 30, 1985. Although OBEMLA did not require final evaluation reports to be submitted until December, 1985, those reports that were submitted to the projects in September and October afforded the most opportunity for being used for planning and decision making for the next grant year especially since the school year in Florida began in August, 1985. All seven reports (58.3%) which included a submission date were received prior to the December OBEMLA deadline and five of the seven reports (41.7%) arrived in a timely manner for planning and decision making purposes, one in April (process evaluation only), one in August, one in September, and two in October. Two reports arrived in early November. Five reports (41.7%) of the sample of 12 reports received by the meta-evaluation did not indicate a submission date. One of the five reports which did not provide a submission date had not been received by the project by May 8, 1986, but was submitted by the project to the meta-evaluator in time to be received by May 30th. One of the two reports not submitted to the meta-evaluator had not been received by the project by May 8th and was ultimately omitted from the sample. The two Title VII projects which had not received their final evaluation report by May 8, 1986, and the project which requested in writing not to be included in the sample, all employed the same evaluator. Since data were available on the timeliness of only 58.3% of the project final evaluation reports, the standard was not met, even though five of the seven dated reports did arrive in time for planning and decision making.

A8--Evaluation impact. Impact means the utilization of evaluation results for decision making and program planning; the influence the evaluation results exert on the project that was evaluated. In most cases evaluation results that are submitted without recommendations by the evaluator have little impact on program planning and decision making. Five reports (41.7%) did not include evaluator recommendations. None of the final evaluation reports addressed plans for the evaluator to act as change agent or specific steps for evaluation utilization. Standard A8, Evaluation Impact, was not met by the sample.

Feasibility Standards

The following three standards were designed to ensure that the evaluation process would be realistic, practical, diplomatic, and cost effective.

B1--Practical procedures. The activities involved in the evaluation need to be realistic for the context of the evaluation setting and should be carefully planned to avoid disruption to normal schedules, both in the project office and at each site. Since 83.3% of the evaluation reports did not identify the point in the project's functioning at which the evaluator was hired, it is impossible in most cases to determine who was responsible for data gathering procedures, which are potentially the most disruptive evaluation activity. Field notes, not final evaluation reports, are the evaluator's usual repository for detailing procedures and commenting on the execution of each step of the evaluation. Since field notes were not part of the meta-evaluation and no additional data were specifically collected to address this standard, no judgment can be rendered.

B2--Political viability. The early identification of official and informal power structures and special interest groups gives the evaluator the option to make constructive use of political forces to avoid political conflict, attempts to bias, or misapply results. By identifying the right-to-know audiences and giving them the opportunity to express positions and concerns regarding the evaluation the evaluator can demonstrate integrity and give all stakeholders the assurance of an impartial evaluation. As was mentioned in A6, no report listed right-to-know audiences as such, although administrators, teachers, and parents were asked to respond to questionnaires in 83.3% of the projects and were interviewed in 16.7% of the projects. The standard was not met.

B3--Cost effectiveness. Evaluations should be as economical as possible, with the benefit in terms of the value of all the results derived being equal or exceeding the costs which are the total social and monetary value of all the human and physical resources expended in the evaluation effort. This standard addresses the situation where an evaluator negotiates with a client for an evaluation budget. In the Title VII evaluations in Florida, the evaluator was offered a fixed sum for his or her time and expertise which was predetermined in the project proposal budget. The district absorbed such costs as the purchase of testing instruments, test administration by district personnel, test scoring, xeroxing, and the use of district facilities such as district mail. Since the evaluation is a requirement of the grant and the expense of the evaluator is an expected grant budget item, the question of cost effectiveness was not addressed by any of the projects. This standard was not applicable to the Florida 1984-85 Title VII project final evaluation.

Propriety Standards

The next eight standards deal with the legal and ethical conduct of evaluations and respect for the rights of human subjects and those who are affected by the results.

C1--Formal obligation. This standard addressed the written agreement between client and evaluator and recommended that a copy of the evaluation contract be appended to the final evaluation report. No mention of contractual agreements was made in or appended to any of the final evaluation reports. It is clear that this standard was not met, but it is not clear whether this standard is applicable to contractual agreements between school districts and contractors in the State of Florida.

C2--Conflict of Interest. This standard suggested that a conflict of interest in terms of a philosophical or political point of view may be unavoidable, but if disclosed openly and honestly, should not compromise the objectivity or credibility of the report. No such disclosures were made in the sample of 12 final evaluation reports. An objective tone seemed to characterize the majority of evaluation reports (75%) with no evidence of evaluator hostility to bilingual education and only limited evidence of a loss of objectivity (25%) in becoming an apologist for the project (see explanation for C7). There is always potential for a loss of objectivity in internal evaluation but the final reports submitted by district personnel were written in a professionally objective tone. Judgment was reserved on this standard.

C3--Full and frank disclosure. A candid and honest presentation of all relevant evaluation findings including the evaluator's judgments, limitations of the evaluation, and implications for the findings and recommendations are requirements of this standard. As was mentioned in C2, a majority of the evaluations were written in an objective tone, but only four reports (33.3%) addressed limitations and constraints. Seven evaluation reports (58.7%) provided recommendations but none of them discussed implications for the findings and recommendations. The standard was not met in the sample of Title VII final evaluation reports.

C4--Public's right to know. Right-to-know audiences were defined by the Joint Committee as those who are entitled to be informed about the results of the evaluation for the following reasons: they are the client; they commissioned the evaluation; they bear legal responsibility for the project; they funded the project through taxes or gifts of time or money; they supplied data; they are stakeholders such as staff, parents, students, and representatives of the mass media. No evaluation report identified right-to-know audiences. The standard was not met.

C5--Rights of human subjects. The rights and welfare of human subjects must be protected. These rights are derived from the law, ethics, common sense, and courtesy. Violation of legal and ethical rights may subject the evaluator to legal prosecution or professional sanctions. Four final evaluation reports (33.3%) violated the rights of human subjects. In three reports (25%) parent sign-in sheets were included in the appendix as proof of PAC activity. The sign-in sheets contained such personal information as name (signature), school name, address, and phone number. One of these three reports included copies of grade level summary profile sheets, listing the name of the school, aide, grade, year, names of all LEP students, date of entrance into the program, pretest-posttest raw scores for language proficiency tests, and pretest stanines for the achievement test. A fourth final report included a copy of the bilingual follow-up form for each student which had been placed in a mainstream classroom. This form listed the student's name and the mainstream teacher's comments regarding academic and social progress as well as a narrative identifying deficiencies and assessing degree of deficiency. The evaluator could have noted the presence of sign-in sheets on file in the project office, and the last two documents could have been included in the report if all identifications of school, aide, student names, and identifying descriptions had been whited out before copying, although a summary of the relevant information would have been even better. This standard was not met.

C6--Human interactions. This standard addresses the evaluator's respect for human dignity and worth in interpersonal transactions with those who participate in the evaluation. Of particular importance in a Title VII bilingual evaluation is that the evaluator make every effort to understand the social and cultural values of the participants and the appropriate protocol for school visits and staff interviews. Only one piece of data appropriate to this standard was included in the sample of final evaluation reports, a letter to the schools informing them of the evaluation's purpose, introducing the evaluator, advising them of the date the evaluator would be on campus, and indicating who would be observed and/or interviewed. This standard could not be judged as no further data were collected in the meta-evaluation for Standard C6, Human Interactions.

C7--Balanced reporting. A balanced evaluation report provides a fair assessment of both strengths and weaknesses of a project. The report should be complete; where data omissions occur, they should be identified. Weaknesses should not be deleted to avoid embarrassment. This standard is especially relevant to Title VII project reports where there is strong competition for funding and uncertainty about governmental policies toward the field in general and congressional funding in particular. Many districts fear the loss of federal funds to the extent that any indication of a weakness in the project produces anxiety rather than providing direction for improvement. There is a feeling that OBEMLA would prefer a public relations document listing project accomplishments and student achievement that could be aggregated nationally and presented to Congress at refunding time, rather than an assessment of merit and worth, strengths and weaknesses, with recommendations for project improvement. In the sample of 12 final reports, 6 evaluations (50%) achieved a fair measure of balance in addressing strengths and weaknesses, while 50% did not address weaknesses. A balanced report also has to do with the completeness of the data collection. One project with five components presented evaluative data on one component, descriptive data on a second, and completely ignored the rest, except to combine achievement data for all five components together, although they served totally different student populations and the treatment was designed for the specific student population and therefore differed. No project which served more than one language group separated achievement data by language group although this practice of presenting data by language group is recommended by OBEMLA. Finally, in order for a report to be balanced, first-hand knowledge of the project and sites as they actually functioned is a requisite. Only two evaluators (16.7%) reported personal observation data, and only four evaluators (33.3%) reported face-to-face interviews, indicating first hand knowledge of the project as it actually functioned. The standard was not met.

C8--Fiscal responsibility. This standard deals with the evaluation budget, as did B3, Cost Effectiveness. This standard, directed to the evaluator, addresses the maintenance of accurate accounting records of expenditures for the evaluation. This standard is not applicable to Florida Title VII projects because the evaluator is paid for his or her time and expertise but the school district maintains control of the evaluation budget.

Accuracy Standards

These 11 standards were designed to ensure that technically adequate information is provided for the assessment of merit and worth of the project.

D1--Object identification. The project should be described in enough detail that unique features can be identified and associated with program effects. Discrepancies between proposed and actual program characteristics should be noted and special assistance to students should be described. Reported for question 1 were the meta-evaluation results related to program characteristics, project type, size of project, languages, student characteristics, program characteristics, exit and entry criteria, and program emphasis. Although many of the evaluations did an excellent job in describing one characteristic or another, taken as a whole, the standard was not met.

D2--Context analysis. Sufficient contextual variables should be addressed in a final evaluation report so that the reader may judge the effect of the contextual conditions on the project and under what similar conditions the findings may apply. Reported for question 1 were the meta-evaluation results related to the following contextual variables: geographic location, description of district, available resources, community characteristics, and local school context. Although many projects did describe certain elements of the context, the standard was not met.

D3--Described purposes and procedures. The evaluation purposes and procedures should be sufficiently described so they can be identified, replicated, and/or assessed. The evaluation purposes identify objectives and intended uses of results while the procedures indicate how the data were gathered, organized, analyzed, and reported. Question 2 reported the meta-evaluation results related to the description of evaluation purposes and procedures. Six reports (50%) did not address the purpose of the evaluation and certain procedures were addressed in about the same number of reports but no project reported specifically on evaluation procedures in the final evaluation report. Procedures may have been addressed in a separate report of field notes not included in the meta-evaluation. The standard was not met.

D4--Defensible information. Documentation of the sources of information is addressed in this standard so their adequacy may be assessed. Sampling procedures, attrition, copies of project-developed instruments, observation and interview schedules and notes, lists of standardized tests, and rationale for selection of instruments for use with the sample are all considered in this standard. In two reports (16.7%) there was a confusion between random and matched sampling for control groups. In both reports the samples were reported as drawn randomly at one point in the report and then defined as a matched sample in another part of the report. Attrition was severe in several projects but the effects of attrition were not addressed. One project inflated the number of subjects served by adding the individual attendance figures for each week of a two-cycle 16-week voluntary training program where the same subjects returned each week and represented a total attendance of 488 as the number of individual subjects treated. By the end of the first cycle 41 subjects were tested. By the end of the second cycle, 10 subjects remained for testing. Of the subjects who completed the first cycle, 75% had dropped out by the end of the second cycle, but the attrition was not addressed. Another project started with 250 subjects and ended with 161 subjects, without a word acknowledging the effects of attrition on the project and the results. Many projects developed and used criterion referenced tests (CRTs) but did not mention field testing or assessment of the validity and reliability of the instruments. In one report raw data and a frequency count were presented with no explanation of cut-off points or what the numbers indicated. Data sources were directly questioned by one evaluator. Most reports did not describe the instruments used and did not provide a rationale for instrument selection or address the content validity of the instruments selected for the project curriculum. For the reason that many of the sources of information could not be assessed as adequate, this standard must be judged as not met by the sample of 12 Title VII final evaluation reports.

D5--Valid measurement. The intent of this standard was that standardized instruments and other data gathering instruments and evaluation procedures must not only be valid in general but that they must be assessed as sound for the project in which they are used, the curriculum content and specific student characteristics, test administration, scoring, and interpretation of results. The topic of validity was addressed in one form or another in 83.3% of the final reports. Four reports (33.3%) discussed project attempts to validate CRTs and other project developed materials. Two reports (16.7%) mentioned that multiple data sources validated the results, with one of the two reports noting that the wording of questions was monitored so as not to affect the responses. The evaluator questioned the validity of instruments in three reports (25%). Most of the reports, although they mentioned validity, did not fulfill the intent of the standard to assess the validity of the data collection process for the special characteristics of the project.

D6--Reliable measurement. Instruments should be chosen that have acceptable reliability for their intended uses and measurement techniques should be administered in such a way that they minimize unreliability and assure that the information obtained is sufficiently reliable for the intended use. According to the Eighth and Ninth Mental Measurement Yearbooks (Buros, 1978; Mitchell, 1985), many of the instruments available for use in bilingual settings do not have adequate reliability data reported. The intent of this standard was that instruments should be selected which have adequate reliability data, but that the reliability of the use of the instruments should also be a concern. For example, if more than one person administers an oral language dominance test or proficiency test, were they using the same criteria and would they come up with the same judgment? Did they receive training before they administered the test, and was there an assessment of interrater reliability? The reliability of the group mean should also be examined. None of the final evaluation reports discussed interrater reliability, the reliability of the group mean, test-retest reliability, or internal consistency reliability. A few of the reports noted the absence of reliability data for published standardized tests and widely used CRTs. One report presented data which clearly indicated a ceiling effect, where student scores on the pretest were so high that posttest scores were nonsignificant, but the evaluator did not mention that the level of the test was too low for the students tested, therefore yielding unreliable data. Since none of the reports posed the question of how reliable the data collection procedures for the specific project and students were, it must be judged that this standard was not met.

D7--Systematic data control. This standard was designed to institute a systematic program of training, controls, and accuracy checks to monitor and review the data collection, analysis, and reporting procedures to assure that data used would be as error-free as is possible. One of the dangers in a Title VII project is that the bonding between aides and their students might be so strong that there would be temptation to assist students with additional translation during standardized testing. One evaluator questioned if this had not happened with the previous year's standardized achievement scores, which he had planned to use as a comparison. Another problem involves

the whole question of testing in a second language, English. At what point is the test assessing knowledge of English and knowledge of specialized testing vocabulary and skills and at what point is the test assessing knowledge of content? Were test administration procedures standardized in regards to previous instruction in testing vocabulary and manipulation of response sheets and explanations of directions? Have systematic checks for error in collecting, processing, and reporting data been implemented? In numerous reports, columns of data did not add up to the indicated total, probably a clerical error which went unchecked. If clerical error occurred frequently in the final report, which should have been proofread, was there any check for accuracy in keypunched data? One chart reported percentages of students mastering 70% on a CRT. In a fairly long listing of schools, one school had 100% of its LEP students master 70% while another school had 0% of its LEP students master 70%. This could have been an error involving nonstandardization of test administration, incorrect key punching, incorrect data analysis, or a serious problem indicated in the program. Whatever was the cause, all the numbers on the chart added up correctly with the 0% and the evaluator did not address the matter in the narrative. In another report, the evaluator summarized and reported findings which contradicted data reported in tables.
It was reported in one final evaluation that the aides received training previous to test administration and another project staff wrote their grant specifically to provide training to the district and project staff for implementing a language management system for testing

and tracking the progress of individual students. Finally, one evaluator recommended development of a data control system. Although there appears to be an awareness of the need for systematic data control, the standard could not be judged as having been met at the present.

D8--Analysis of quantitative information. The analysis of data in an evaluation should be appropriate and systematic to ensure that the interpretations are supportable and that the evaluation questions can be answered. For question 2 the quantitative data analysis procedures were identified and described for the sample of 12 Title VII final evaluation reports. Two of the most frequently applied analyses were a comparison of the LEP students average performance to the average performance of a similar national norming population on a standardized test and the attainment of mastery of a specific body of information, usually on a project-developed CRT. There are problems with using a standardized test and comparing LEP students with national norming populations. The comparison is only successful if the two groups being compared are similar and most norming populations do not contain large numbers of limited-English proficient students, a large percentage of students who have not been raised in the American culture, and a disproportionate number of students in the lower socioeconomic brackets. Two projects used district norms in addition to national norms to have a more satisfactory comparison group. Regression to the mean is also a problem since students are selected for Title VII projects because of lack of success in the regular classroom and lack of comprehension of the English language. Another problem found in the sample was that several projects used the same test to identify the student for the project and as a pretest. In all three of these conditions selection for low achievement, limited English, and identifying students for project with the pretest, the gain indicated cannot be attributed to the success of the project but to a statistical artifact.

The use of CRTs, especially project-developed CRTs, is also problematic in that seldom are the tests field-tested and validity and reliability data are seldom collected. Several projects relied heavily or exclusively on CRTs. Information was provided on how the LEP students performed in relation to a particular cut-off score but there was no information as to how the project's students compared with other similar LEP students in other projects or whether the gap between LEP performance and the performance of a norming population had widened, remained the same, or had been narrowed.

There were specific problems with the data analysis procedures in several reports. One error involved the calculation of a group mean by averaging the means of the schools or classes listed rather than calculating the group mean from the raw data. Increases in posttest scores were reported without a test of significance and gain scores were presented as the percentage of students who increased their posttest over their pretest. Taking into consideration the data presented above and in response to question 2, it must be judged that the standard was not met.

D9--Analysis of qualitative information. Nonnumerical data, in the form of documentation, observations of the project sites in action, and the collection of the opinions of the participants, staff, administrators, parents, and evaluation audiences should be appropriately and systematically analyzed to ensure supportable interpretations. Triangulation of data such as the use of questionnaires, observation, interviews, documentation, and results from tests to answer the same question strengthens confidence in the accuracy of the findings and usually provides an explanation for the findings. Presented in relation to question 2 was the frequency of the use of questionnaires, documentation, interviews, and observation in the sample of 12 final evaluation reports. The major problem that can be seen from the analysis is simply underuse of qualitative information. Direct observation and interview would be indicated in situations, for example, where attrition is very high or where one site has 100% of subjects mastering a criterion and another site has no subjects mastering the criterion. Qualitative information derived from documentation and observation is essential in establishing that a program has been implemented as proposed. Interviewing and observing staff who have participated in inservice training is essential in judging whether the training had an impact on classroom performance. Merely counting frequencies of inservice training or number of contacts of a social worker or counselor does not answer the question of whether training or contacts were effective. One problem that was evident in the construction of questionnaires was the tendency to lead the response to the desired answer by the wording of the question, invalidating the responses. One project report presented the headings for a Likert-type scale out of order so that responses were confusing. This may have been a clerical error or an accurate representation of the misuse of the scale but the end result was that no confidence could be placed in the responses from the survey. Although a few project reports triangulated data, 83.3% of the reports were written without first-hand knowledge of the project that was evaluated. The standard must be judged as not met for the reason of infrequency of use of qualitative data.

D10--Justified conclusions. The conclusions, both judgments and recommendations, must be based on sound logic and appropriate data, and reported with an account of procedures, underlying assumptions, and possible alternative explanations of the findings and reasons for rejection. The conclusions should answer the evaluation questions. Limitations and cautions in interpreting results and possible side effects should be discussed. The designers of the standard warned against basing conclusion on a single source of information, a single type of data, or a single analytic technique. Only two project reports (16.7%) formalized evaluation questions and an additional eight reports (66.7%) substituted project objectives for questions. In relation to question 2, information was provided that only seven reports (58.3%) concluded with evaluator recommendations. In the 10 reports (83.3%) which formulated evaluation questions or objectives, the recommendations and/or conclusions addressed the questions or objectives. Alternative explanations were presented in only one report (8.3%), while limitations and cautions in interpretation were presented in four reports (33.3%). The appropriateness of the data has been discussed previously, as has the tone of the report, objective (75%) or apologetic (25%). The standard was approached but not met.

D11-Objective reporting. The 30th and final standard was designed to safeguard the evaluation findings and reports from evaluator bias and to assure that the results were based on impartially gathered facts and presented in an unbiased report. Reports may be considered biased by omission if they present only one value interpretation or biased by commission if there is a deliberate design to cover-up, no matter how good the intentions or how worthy or needy the subjects are of treatment. Evaluator ignorance or carelessness and project pressure can also contribute to a biased report. This standard should be considered in relation to Standard A4, Valuational Interpretation; C2, Conflict of Interest; C3, Full and Frank Disclosure; and to C7, Balanced Reporting. The following information was gleaned from these standards: 75% of the reports were written in an objective tone; 50% of the reports considered the value judgments of the project without reference to additional audiences; 50% of the reports addressed both weaknesses and strengths of the project, but 50% only addressed project strengths; 58.3% of the evaluation reports expressed evaluator recommendations, but 41.7% did not present recommendations; 83.3% of the reports were written without mention of the evaluator observing the program in action; 66.7% were written without interviewing project staff, mainstream teachers, administration, students, or parents; limitations and cautions to interpreting results were addressed in 33.3%, and not addressed in 66.7%; and, finally, a case was built for the success of a program on results that were presented as not significant in the data display table but were labeled as significant in the conclusion. What this indicates is bias by omission in the majority of the instances and one case of bias by commission or evaluator carelessness. Standard D11, Objective Reporting, has not been met.

Summary of Results for Question 3

Question 3 asked, "do Title VII evaluation reports meet the current accepted professional standards of good evaluation practices?" The answer is quite simply, "no." Although some of the individual reports met some of the standards, taken as a whole, the sample of 12 final evaluation reports did not meet any of the Joint Committee's 30 professional standards of good evaluation practices.

Question 4: Discrepancies Between Proposal and Report

One requirement of the application proposal for acquiring a Title VII federal bilingual education basic grant is that the proposal must contain an evaluation plan. The U.S. Department of Education, Office of Bilingual Education and Minority Language Affairs (OBEMLA) publishes the current legal requirements and selection criteria for the evaluation plan in each application packet. Authority for the 1984-85 bilingual program was contained in Section 721 of the Elementary and Secondary Education Act of 1965, as amended by Educational Amendments of 1978 (Public Law 95-561). Specific legal requirements for the 1984-85 proposal evaluation plan were stated in the 1982 Code of Federal Regulations (CFR) Title 34, Section 501.23 and the selection criteria were stated in 34CFR 501.30(e). The federal legislation which addressed regulations for evaluation plans has been amended twice since the 1978 amendments.

Various aspects of the 1984-85 Title VII bilingual education final evaluation reports have been examined in questions 1, 2, and 3, including an analysis of Florida final reports in relation to professional standards of good evaluation practices. In this question, 1984-85 proposal evaluation plans were examined in relation to their corresponding final evaluation reports. Question 4 asked, "are there major discrepancies between LEA evaluation proposals and summative evaluation reports? What is the nature of these discrepancies?" Data which addressed question 4 are presented in the following order: proposal evaluation plans; discrepancies between plans and reports.

Proposal Evaluation Plans

Description of sample. Title VII grants are awarded for a period of up to three years with a noncompeting continuation reapplication required for the second and third years. The reapplication proposal is generally less comprehensive in coverage than the first year application proposal. Eleven 1984-85 proposals were submitted to the meta-evaluator corresponding to the sample of 12 final evaluation reports. Four of the 11 proposals were first year applications and 7 were noncompeting continuation reapplications. Three of the noncompeting continuation reapplications contained no evaluation plan, establishing the sample of proposal evaluation plans at eight, four new proposal evaluation plans and four noncompeting continuation reapplications.

Federal requirements. Just as the Joint Committee Standards were used to judge the adequacy of the final evaluation reports, the applicable federal regulations which governed the 1984-85 proposal application are the standards by which the proposal evaluation plans must be judged. Table 1 presents a listing of the 16 federal evaluation requirements for proposal evaluation plans and the number of requirements met by the four new proposals and the four continuation proposals.

The new proposals as a group achieved 75.6% of the requirements of the federal regulations for evaluation proposals for the 1984-85 application. Three continuation proposals did not include an evaluation plan at all and were excluded from the sample. When considered as a group, only 42.2% of the federal requirements were met by the four continuation proposals which included evaluation plans in their proposals. The quality of the individual continuation proposals was quite similar but there was noticeable differentiation in the quality of the new proposals. Two new proposals both met 14 of the 16 requirements (87.5%) while the other two new proposals met 11 (68.8%) and 9 (56.2%) of the requirements. All of the proposals were funded, however, since in 1984-85 the evaluation plan was worth only 15 points out of a total of 110 points. No project could be funded that received less than 60 points in the 1984-85 grant awards, so that if a project lost all 15 points for the evaluation plan, as did the three proposals that were excluded from the sample, the project could still be funded.

Discrepancies Between Plans and Reports

In the Discrepancy Evaluation Model, a comparison between the standard (S), what should be, and the performance measure (P), the actual characteristics of the object to be evaluated is called the

Table 1

Evaluation Requirements Met by New and Continuation Proposals

Evaluation Requirements	New Proposals		Continuation		Total	
	Yes	No	Yes	No	Yes	No
Evaluation of Progress	4	0	4*	0	8*	0
Evaluation of Achievement	4	0	4*	0	8*	0
Instructional Objectives	4	0	3	1	7	1
English Language Skills	4	0	4*	0	8*	0
Utilization of Results	3	1	2	2	5	3
Evaluation Design	2	2	0	4	2	6
Attainment of Objectives	4	0	2	2	6	2
Evaluation Instruments	4	0	3	1	7	1
Data Collection Methods	3	1	1	3	4	4
Data Analysis Procedures	2	2	0	4	2	6
Time Schedules	4	0	1	3	5	3
Staff Responsibilities	4	0	2	2	6	2
Comparison--Absence of Project	2	2	0	4	2	6
Comparison--Historical/Stat.	2	**	1	3	3	3
Sampling Procedures	2	**	**	1	2	1
Evaluation Questions	1	3	0	4	1	7
Totals Achieved	49	11	27	34	76	45
Percent Achieved	(75.6%)		(42.2%)		(59.4%)	
Combined YES NO Total Possible	64		64		128	

* Unspecified, undeveloped criterion referenced tests

** Not applicable in all designs

discrepancy (D). Judgments of worth or adequacy are based on the discrepancy between the standard and the performance measure, D=S-P. In the case of the proposal evaluation plan, when taken as a total sample of eight plans (P), only 59.4% (D) of the federal regulations (S) were met. In the case of the final evaluation report, 0% (D) of the Joint Committee Standards for Educational Evaluation (S) were met by the sample of 12 final evaluation reports (P).

Question 4 asked if there were major discrepancies between the LEA proposal evaluation plan and the final evaluation report. In determining a discrepancy, in this case the difference (D) between the proposal evaluation plan and the final evaluation report, a standard (S) must be determined. The proposal must be considered a flawed standard, since the federal regulations, which have since been amended twice, lack specificity and professional rigor, and yet are only 59.4% achieved by the sample of eight plans. The final evaluation reports (P), although not meeting the standards of the Joint Committee appeared to be less flawed than the proposals. Consequently, the formal Discrepancy Evaluation Model was abandoned in favor of a descriptive analysis of differences between the proposal and the final evaluation report.

Description of differences--new information. The budget, of prime importance in the Title VII proposal, was rarely addressed by the external evaluator in the final evaluation report. Only one final report (8.3%) of the sample of 12 included a negotiated budget as part of the implementation evaluation. The following information was available from the budget: total amount requested for funding; the exact number and type of staff positions requested; staff development allocations, including college courses, consultants, and the number of aides requested for classroom duty per inservice day; amount for materials acquisition and development; and the amount set aside for the external evaluator. This last amount, external evaluator, ranged from $2,000 to $5,850, with a total of 8 of the 11 proposals reporting an amount. The mean amount budgeted for the external evaluator was $3,806.25. The median amount was $3,500, and the mode was $3,250 budgeted for the external evaluator. No relation appeared to exist between budgeted amount for evaluator and total size of budget, number of students served, new or continuation proposal, or quality of final evaluation report.

The proposal budget is not the final budget for a project. Once the grant is awarded a negotiation process between the project and OBEMLA finalizes the negotiated amount that is awarded. In the case of the one report which furnished the negotiated budget, the proposed amount dropped from $166,443 to a negotiated budget of $105,147. From this example of one, no conclusions can be drawn, except that the negotiated budget should be used for analysis of budgeted items.

Description of differences--more proposal information. Title VII funds are awarded on a competitive application basis. Fifteen points were awarded in the 1984-85 application for need, 10 points for rationale of project sites and participants, and 20 points for commitment and capacity, totaling 45 out of the necessary 60 points for topics that describe context and process variables. It is no wonder, than, that the proposals, written by the project staff and rewarded for descriptive information as a whole included more descriptive information than the final evaluation reports.

The proposals as a whole included more information about the following topics than did the final evaluation reports: geographic location, description of district, available resources, community charactersitics, local school context, the languages of the subjects, program characteristics, materials, curriculum, and plans for parent participation/community involvement.

Description of difference--more final report information. The final evaluation report gave a more accurate picture of the actual number of students who received treatment but were in general not as complete in descriptive information as the proposal.

Differences between evaluation plan and the final report. Analysis of the proposal evaluation plan, Table 1, indicated that the proposals were weakest in the following areas: formulating evaluation questions, evaluation designs, comparison procedures to estimate what the performance of participants would have been in the absence of the project, data analysis procedures, and data collection methods. According to the analysis of final evaluation reports for question 2, these very same problems appeared to be weaknesses in the final reports. Only one proposal formulated evaluation questions and only two final evaluations reported the use of evaluation questions. Analysis was simply not addressed in most proposals and technically inadequate and very simplistic in the final evaluation reports.

Evaluation models were not required by federal regulations to be specified in the proposal evaluation plans. Some of the proposals tentatively addressed the specifications of a model but confusion of terminology was more prevalent than clear presentation of an evaluation model in the proposals.

Summary of Results for Question 4

Question 4 asked, "are there major discrepancies between LEA proposals and summative evaluation reports? What is the nature of these discrepancies?" The answer is "yes," there are discrepancies. The evaluation plan and the proposed budget were both included in the proposal, but only one final evaluation report included the negotiated budget. Proposals included more descriptive information on context, program characteristics, and process variables than did the final evaluation reports. The final evaluation reports, however, gave a more accurate accounting of the number of students who received treatment. Significant similarities were also noted. The weakest components of the evaluation plans also proved to be weaknesses in the final evaluation reports, such as data analysis, evaluation questions, designs, comparison procedures, data collection, and an unclear conceptualization of an evaluation model.

Conclusions

Answers to the four main research questions, based on the meta-evaluation of ESEA Title VII bilingual education project evaluations, have comprised this chapter.

Context, program characteristics, and process variables were generally under-represented in the final evaluation reports according to the results of question 1. In order to assess whether all of the proposed aspects of a new project are functioning, an implementation evaluation should be completed first to assure that the project is set in place and is evaluable. More attention to questions of implementation need to be addressed in Title VII project evaluations, as well as more attention to the description of the context of the project, specific characteristics of the program, and process variables.

Evaluation models and designs, according to the results of question 2, need to be defined and employed consistently. More attention needs to be placed on collecting and analyzing a wide variety of valid and reliable data, including data from personal observation of the actual functioning of the sites in the project. Data analysis procedures need to be strengthened, with attention to technical adequacy. The evaluators need to be identified by name, title, address, and to establish their credibility by the inclusion in the final report of a statement of qualifications and experience. The final reports should adequately describe the context, program characteristics, and process variables of the project as well as the specific characteristics of the curriculum, so that the treatment could be replicated in other sites and districts if the results indicate success. Sound and specific recommendations need to be provided by the evaluator for project improvement.

The results for question 3 indicate that none of the 30 professional standards for good evaluation practices were met by the sample of final evaluation reports. One cannot infer from such results that evaluation practices in Florida are any worse than they are in any other state, since no other state has had such a thorough examination of evaluation practices, with the possible exception of California where the meta-evaluation of a sample of Title VII final evaluation reports was conducted by Paula Martin in 1981. The assessment of final evaluation reports against professional standards offers Florida Title VII project administrators a very specific goal to work toward in improving evaluation practices. Title VII project administrators now have available to them a basic explanation of what should be required of an evaluator in terms of credentials for hiring and criteria for planning and judging the quality of the evaluation products. The 30 Joint Committee Standards should be held up as 30 goals to work toward in improving evaluation practices in the State of Florida and in all other states and territories which have bilingual programs which receive federal funds.

Based on the results of question 4, it is evident that the writers of application proposal evaluation plans could well benefit from employing evaluators trained in the Joint Standards to assist in the construction of a technically adequate evaluation plan. The current practice is to employ the evaluator after much of the data have been collected, with the expectation that a professionally acceptable analysis and final evaluation report will result without the evaluator's input into the proposal decision-making stage.

The final chapter, Chapter V, contains an overview of the study, summary of methodology, summary of results, discussion of findings, and implications for the fields of evaluation and bilingual education. Chapter V ends with general recommendations and recommendations for further research.

CHAPTER V
SUMMARY, DISCUSSION, IMPLICATIONS, AND RECOMMENDATIONS

This chapter provides a brief overview of the meta-evaluation of ESEA Title VII bilingual education project evaluations, followed by a summary of methodology, and a discussion of findings. Implications and recommendations for further research are also included in the chapter.

Overview of the Study

This researcher investigated the current status of evaluation in the field of ESEA Title VII bilingual education in the State of Florida for fiscal year 1984-85.

Rationale

Criticisms of bilingual education at the local education agency (LEA) project level can neither be validated nor refuted until there is a thorough evaluation of the most recent bilingual final evaluation reports, hence the need for meta-evaluation. According to Stufflebeam (1974a, pp. 70-71) meta-evaluation should describe and judge the worth and merit of the final evaluation report and should provide recommendations for improvement and utilizations of evaluations.

Not only must bilingual education provide valid and reliable data to counter the attacks of critics but also, as Fuentes (1986) indicated, the field must provide empirically-grounded information on the effectiveness of ongoing practices and on the need for improvement of practices to guide planners in designing the most effective programs. Sanders (1981) identified this type of data as his conceptual analysis stage, the first stage for a theoretical design of a national system for monitoring federally funded bilingual programs. Unfortunately, the collection of such contextually rich, descriptive data of project characteristics and practices as implemented has not been addressed in OBEMLA's current three-year contract with SRA Technologies, 1985-1988, to develop and field test an evaluation system for bilingual education.

Purpose

The purpose of this study was to conduct a meta-evaluation of Florida Title VII LEA bilingual education project evaluations, fiscal year 1984-85, to provide information for accountability of evaluation work. This study examined evaluation model and design characteristics, data collection procedures, testing instruments, data analysis techniques, and general evaluation methodologies, as well as discrepancies between proposed evaluation plans and summative evaluation reports.

Summary of Methodology

The sample consisted of 12 final evaluation reports and 11 corresponding application proposal evaluation plans voluntarily contributed from the population of 14 Title VII Bilingual Education Basic Grants that were funded in the State of Florida, fiscal year 1984-85. Four meta-evaluation instruments developed by Paula Martin (1981) were employed in the initial data gathering stage. Data collection ranged from a simple check on a checklist type instrument indicating whether specific items were or were not addressed in each proposal and final report to copious notes describing and rating the quality of the primary evaluator's report in terms of such factors as comprehensiveness of coverage, timeliness, and technical adequacy.

Analysis of the meta-evaluation data employed descriptive statistics to characterize various aspects of the sample of evaluation designs and reports. Frequency counts, means, modes, and percentages were calculated where appropriate to identify the most common characteristics in the designs and reports. Averages were used to determine the extent to which various context and process variables were developed and objectives implemented. Discrepancy analysis was employed to identify the differences between the evaluation proposal and the final evaluation report.

Of major concern to the meta-evaluator was the selection of appropriate criteria for judging the meta-evaluation. The final evaluation reports were analyzed in relation to the 30 standards of good evaluation practices established by the Joint Committee on Standards for Educational Evaluation (1981). The application proposal evaluation plans were analyzed in relation to the federal regulations for FY 1984 as stated in the application package (U.S. Department of Education, ED Form 4561, 1983).

Discussion of Findings

The meta-evaluation findings are discussed in the order of each evaluation question.

Question 1: Which context and process variables are addressed in Title VII project summative evaluation reports within the State of Florida?

Sufficient contextual variables should be addressed in a final evaluation report so that the reader may judge the effect of the contextual conditions on the project and under what similar conditions the findings may be applicable. Unfortunately, as the results in Chapter IV indicate, the reader would not be able to make such a judgment from reading the final evaluation reports in the sample. Only 16.7% of the reports identified the geographic location of the project in the narrative and only 25% of the reports included a brief description of the district. In about half of the reports the reader was able to determine the district's contributions in terms of resources to the project and why and how the specific sites were selected. The Title VII sites were identified by name or number in only 58.3% of the reports and there was some confusion as to the exact number of sites in several reports. The one contextual variable that was addressed most frequently was the ethnicity of the community, mentioned 83.3% of the time. Inclusion of such student information as the length of time in the U.S., number of days in attendance, and number of years in a bilingual program, and socioeconomic factors of the community would enrich the applicability of the findings of the project.

Program characteristics should provide a picture of the unique features of the project but, again, the findings indicated an incomplete description, if addressed at all. The length of time the project had been in operation was not identified by 25% of the projects, while 58.3% presented an unclear or confusing picture of either or both the number of students served and the number of project sites. Description of treatment was very weak. Because curriculum characteristics, teaching methodologies, language of instruction, and groupings were simply not addressed in the majority of reports programs cannot be categorized as full-bilingual, partial-bilingual, or intensive-English. Data were unclear as to the number of staff or their qualifications or language proficiencies. A description of entry and exit criteria were missing from 66.7% of the reports. The one area that was fairly well documented was the language background of the students in the project but exact numbers of students by Lau categories was not generally addressed. Program characteristics were not adequately described in the sample of 12 final evaluation reports.

Finally, process variables should describe the implementation of the project goals, management of the project, timeline for activities, staff development, materials, progress records, and community involvement. Staff development was the only variable that was adequately addressed but there was little assessment (16.7%) of the staff's application of acquired knowledge to the classroom situation. Only 16.7% of the project reports included an implementation evaluation, documenting that the program had been sufficiently implemented to warrant a full evaluation. These were the only reports that included site visits. The remaining 83.3% of the evaluation reports were based on the assumption that the project must be functioning according to the specifications of the proposals. Only one of the implementation evaluations documented that materials had arrived in good time to be used through the project year but there was no record of the perceptions of the teachers and aides as to the adequacy and sufficiency of the materials available in each project. Curriculum development, revision, or adaptation was discussed in 41.7% of the project reports, but only two reports provided any description of the curriculum at all. Since Title VII requirements address parent participation, the expectancy was that some aspect of the parent participation would be mentioned in most of the reports but 41.7% of the reports did not address this subject. The assessment again must be that process variables were not adequately addressed in the sample of 12 final evaluation reports.

Question 2: What are the characteristics of Florida Title VII evaluation models, designs, and reports in terms of information coverage, content, and procedures?

The label "model" identifies a philosophically inspired conceptualization of evaluation and the methodologies, procedures, tasks, and roles which characterize that conceptualization. The process-product model was identified most frequently (41.7%), although in dialect form (progress-product) in several instances. The formative-summative model was the next most frequently identified, in 16.7% of the reports. Examination of the reports identified an absence of evaluation questions in 83.3% of the reports. The majority of reports, 66.7%, organized the presentation of evaluation results by statement of project objectives, or by instruments, 66.7% of the time. Since several evaluation questions could be constructed for each objective or instrument, the collection of data lacked direction and meaning. It must be judged that the procedures used in evaluating Title VII projects in the sample could not be described as motivated by a thorough understanding and application of a consistent model conceptualization.

An evaluation design provides the plan and organization for the collection of evaluation data. This may be the weakest area found in the meta-evaluation. Randomization, the requisite for the experimental design, was not accomplished in any of the 12 projects. One project did satisfy the requirements for a non-equivalent control group design, but wrongly treated it as a true experimental design. The most frequently used comparison was the norm referenced model. Comparing the performance of an experimental group to the performance of the norming sample is technically adequate only when the two groups can be proven to be similar. The performance of students who are classified as limited English proficient, who are frequently living in homes classified as having low socioeconomic status, and who respond to cultural stimuli which differ sometimes dramatically from the Anglo community, cannot be considered technically adequate when compared with the performance of a national norming sample. One does not need a statistician to recognize the differences inherent in the two groups and hence the inadequacies of the norm referenced comparison when applied to the performance of Title VII bilingual project students, and yet 66.7% of the projects in the sample relied partially or in total on the norm referenced model. Regression to the mean must be considered a factor in any such comparisons where one group is selected for low achievement and lack of English. Comparison of the progress of LEP students with a predetermined mastery level was employed partially or in total in 75% of the projects using project developed and published criterion-referenced tests (CRTs). The validity and reliability of the CRTs employed by the sample either had not been determined, was not addressed, or was questionable. None of the evaluations reported the use of sophisticated statistical analyses. Significance was tested in six of the seven final evaluations which reported pretest-posttest means but the standard deviation was reported in only three evaluations. Three of the six projects which reported percentage data used percentages as their sole statistical analysis. As was reported in Question 4 (D8), numerous errors were found in calculations, charts, and in the improper use of analytical procedures. No confidence can be placed in the reported results of the assessment of achievement as conducted in the sample of Title VII projects.

Although achievement data collected by tests accounted for the majority of the data collected, results from questionnaires were reported in 75% of the reports, interviews were reported in 33.3%, while the direct observation of actual functioning of the program in its various classroom sites was reported in only 16.7% of the reports. Personal, first-hand knowledge of the functioning of the project in its various sites was not mentioned by the evaluator in 83.3% of the final reports, which detracted from the credibility of the reports.

The final part of Question 2 addressed characteristics of the final evaluation reports. The evaluators did not attempt to establish their credibility and in 25% of the reports did not identify themselves by name. It was not always stated whether the evaluator was an internal or external evaluator. In the majority of instances no information was provided as to the evaluator's point of entry into the evaluation or when the final evaluation was submitted. Abstracts or summaries of results were absent in most cases, and an informative introduction describing context, program, and process variables were absent in 33.3% of the reports. Evaluator recommendations were absent in 41.7% of the reports and intended audiences were not identified in 83.3% of the final evaluations. The very mechanics of preparing the final evaluation report appeared to have been flawed in the sample of 12 Title VII final evaluation reports.

Question 3: Do Florida Title VII evaluation reports meet the current accepted professional standards of good evaluation practices, as detailed by the Joint Committee on Standards for Educational Evaluation (1981), Standards for Evaluations of Educational Programs, Projects, and Materials?

The explicit intent of this third question was to ascertain the quality of the sample of Title VII final evaluation reports in relation to each of the 30 professional standards of good evaluation practices. Some of the individual reports met some of the standards, but taken as a whole the sample of reports did not meet any of the professional standards of good evaluation practices. Keith Baker (1984), U.S. Department of Education, Office of Planning, Budget, and Evaluation, accused bilingual education researchers of allowing their "ideological needs and political expediency" to compromise the results of their research with "methodological chicanery" (p. 1). Baker's indictment does not seem to fit the results gathered in this meta-evaluation. Rather than "chicanery," the results seemed to indicate a lack of sophistication and a lack of understanding of what was required in conducting a technically adequate program evaluation. As was noted earlier in Chapter I, Tallmadge, Lam, and Camarena (1985a) posited that a lack of guidance, lack of evaluation expertise at the local level, low priority and low funding levels, and technical difficulties inherent in evaluating bilingual programs combined to produce "basically useless data" (p. viii).

The results of the meta-evaluation have not addressed the quality of ESEA Title VII Bilingual Education Projects in the State of Florida. The meta-evaluation addressed the quality of the final evaluation reports of the projects. It is the reports that do not meet professional standards, not the projects. What this does indicate is that Florida Title VII projects now have 30 very specific goals to work toward in improving evaluation practices.

Question 4: Are there major discrepancies between LEA proposals and summative evaluation reports? What is the nature of these discrepancies?

Just as the final reports were judged in relation to the Joint Committee standards, the application proposal evaluation plans were judged in relation to the applicable federal regulations which governed the 1984-85 application for Title VII basic grants. The new proposals, which tended to be more complete, met an average of 75% of the requirements of the federal regulations for evaluation proposals for the 1984-85 application. Three of the sample of 11 noncompeting continuation reapplications contained no evaluation plan and were dropped from the sample. The remaining four continuation proposal evaluation plans met an average of 42.2% of the federal regulations. As was explained in question 4, those projects which selected to eliminate the evaluation plan from their proposal application lost only a total of 15 points out of a possible 110, with a minimum of 60 points required for funding. As further evidence of the lack of importance placed on the evaluation plan by the federal government, in the 1986 application guidelines the evaluation plan is worth a total of 8 points out of a possible 100, a drop in worth from 14% of the total points to 8% of the total points.

There were several discrepancies between the proposal evaluation plans and the final evaluation reports. The final evaluation reports gave a more accurate count of students who actually received treatment. The proposal evaluation plans, however, provided significantly more descriptive information on contextual, program, and process variables and included the proposed budget.

The proposed budget provided such information as the total amount requested for funding, the exact number and type of staff positions requested, staff development allocations, materials acquisition allocations, and the amount set aside for the evaluator. Only one final evaluation included a copy of the negotiated budget, which went from the proposed $166,440 to the negotiated $105,147, a drop of 36.8% in budget. Since information was available on only one negotiated budget no speculation can be made as to trend. The amount requested for the external evaluator ranged from $2,000 to $5,800, with the mean at $3,850, the median at $3,500, and the mode at $3,250. There appeared, however, to be no relation between requested amount for external evaluator and the total size of the budget, number of students served, new or continuation proposal, or the quality of the final evaluation report.

The proposals were weakest in formulating evaluation questions, evaluation designs, comparison procedures to estimate what the performance of participants would have been in the absence of the project, data analysis procedures, and data collection methods, all areas of weakness in the final evaluation reports. As a generalization, when the proposal evaluation plan was well written and met a high percentage of the federal requirements, the corresponding final evaluation was of higher quality than the average. The reverse was not necessarily true, however, that where the final reports were of higher quality, the proposals met a larger percentage of federal requirements. This seems to indicate that technical assistance at the stage of the writing of the proposal evaluation plan might have a positive effect on the quality of the final evaluation reports.

Implications for the Field of Evaluation

Evaluation has developed over the last 20 years from a situation where college professors from the fields of education, measurement, research, and psychology were pressed into service by a federal mandate for evaluation of federally-funded programs to the establishment of an independent profession. In the 1980s graduate training programs in evaluation have been established at recognized universities; evaluation journals are published and line the shelves of research libraries; and professional societies for evaluators have enjoyed wide acceptance in the research community. The profession has established a set of standards by which the quality of evaluations may be judged in order to guard against or deal with malpractice or services not in the public interest and to create increased understanding of the ethics and practice of the profession.

In the case of bilingual education evaluation, just as in the case of medical evaluation, there is a valid ethnocentrism which says that the field is specialized and evaluators must know the field as well as evaluation methodology for the evaluations to be of value. In the absence of evaluators who know the field of bilingual education, Title VII administrators often select professionals in the field of bilingual education to serve as external evaluators rather than professionally trained evaluators who do not understand the complications and intricacies of the field. The future prospects for adding to the ranks of the professionally trained bilingual education evaluators is limited.

Funds for the evaluation of programs are set in the evaluation proposal budget and must go through the negotiation phase with OBEMLA. The average funding level for evaluation of a Title VII project in Florida is well below what an evaluator could expect to make on evaluation contracts in other fields. The low budget discourages highly trained evaluators from obtaining the special initiation to the field that is perceived by Title VII administrators to be a requisite.

The results of the meta-evaluation indicated that the quality of evaluation of bilingual programs is in immediate need of improvement starting with adherence to the standards of the profession. If professional evaluators cannot be easily infused into the field, and if Congress is not likely to increase funding for bilingual projects to improve the evaluation component, the only other viable option, then, appears to be to provide graduate or inservice training in evaluation and to offer the support services of the profession to those who are currently evaluating bilingual programs. The question of licensing evaluators to practice goes beyond the confines of this research project.

Implications for the Field of Bilingual Education

If bilingual education as a field is ever going to be considered a powerful force by the federal government, technically adequate achievement data that can be aggregated nationwide are required. The data that were collected in the sample of 12 Title VII final evaluation reports did not meet Joint Committee standards and little confidence could be placed in the accuracy of the results. What is needed in bilingual education is methodologically sound evaluation plans and evaluations that adhere to the standards of good practice for the evaluation profession. If the federal government does not mandate rigorous evaluations performed by trained evaluators, possibly the field itself should consider the establishment of policy in regards to the evaluation of bilingual programs. Technically adequate evaluations of bilingual programs are a requisite for the "proof" of effectiveness of bilingual instruction, the evidence that U.S. Department of Education Secretary Bennett (1985) sought to validate that the limited-English proficient children in American schools have been and are being successful in becoming proficient in the English language through the intervention of bilingual education.

Recommendations

In consideration of the results obtained from the meta-evaluation of Florida ESEA Title VII Bilingual Projects, the following recommendations are presented for consideration:

1. The replication of this Title VII meta-evaluation research is recommended certainly at the regional level if not at the national level to provide OBEMLA with a data base which will describe the current status of evaluation practices in the field of bilingual education.

2. The establishment of a network of federally-funded evaluation assistance centers, with at least one center for each state, is recommended to provide timely on-site consultation during the planning and writing of the application proposal evaluation plan, to provide training to enable project administrators to become more sophisticated consumers of evaluation services, and to help administrators build into their proposals specific plans for the utilization of evaluation results. It is possible that such services could be coordinated with the already

established Chapter One Regional Evaluation Technical Assistance Centers.

3. The provision of federal funds is recommended for inservice advanced training in evaluation methodology with attention to the Joint Committee standards (1981) from the _Standards for Evaluation of Educational Programs, Projects, and Materials_ for those who are currently evaluating Title VII bilingual projects.

4. The alignment of federal regulations for the evaluation of Title VII bilingual projects with the Joint Committee standards from the _Standards for Evaluation of Educational Programs, Projects, and Materials_ is recommended.

5. The inclusion of the requirement of a thorough implementation evaluation at the conclusion of the first year of the grant is recommended.

6. The preparation of a set of technically adequate handbooks and manuals is recommended. These materials should be written in layman's terms in step-by-step progression which addresses the special problems in evaluating bilingual programs, such as the technically appropriate way of evaluating the effectiveness of a bilingual program with a high percentage of students who move frequently but are not receiving services from migrant programs or are not in the migrant computer system.

Recommendations for Further Research

1. The meta-evaluation of ESEA Title VII bilingual project final evaluation reports should be replicated using a team of researchers so that inter-rater reliability could be tested.

2. Questionnaire and interview components should be included in the meta-evaluation, where project administration, project staff, and evaluators are included in the sample.

3. An evaluation utilization analysis component should be included in the meta-evaluation research project.

4. The four Martin (1981) meta-evaluation instruments, designed specifically for use in bilingual projects, should be refined and their validity and reliability should be established.

5. A reliable and valid computer program for use in the meta-evaluation of Title VII bilingual programs should be developed.

6. The meta-evaluation of ESEA Title VII bilingual final evaluation reports should be replicated using a combination of a meta-evaluation and a meta-analysis of the same sample of Title VII bilingual projects.

APPENDIX A
STANDARDS FOR EVALUATIONS OF EDUCATIONAL PROGRAMS

Utility Standards

A1	Audience identification
A2	Evaluator credibility
A3	Information scope and selection
A4	Valuational interpretation
A5	Report clarity
A6	Report dissemination
A7	Report timeliness
A8	Evaluation impact

Feasibility Standards

B1	Practical procedures
B2	Political viability
B3	Cost effectiveness

Propriety Standards

C1	Formal obligation
C2	Conflict of interest
C3	Full and frank disclosure
C4	Public's right to know
C5	Rights of human subjects

C6 Human interactions

C7 Balanced reporting

C8 Fiscal responsibility

Accuracy Standards

D1 Object identification

D2 Context analysis

D3 Described purposes and procedures

D4 Defensible information

D5 Valid measurement

D6 Reliable measurement

D7 Systematic data control

D8 Analysis of quantitative information

D9 Analysis of qualitative information

D10 Justified conclusions

D11 Objective reporting

(Joint Committee on Standards
for Educational Evaluation, 1981)

APPENDIX B
INSTRUMENTS

The following instruments are available only under separate copyrights:

Instrument #1, the "CES Meta-Evaluation Checklist" (Martin, 1981);

Instrument #2, "Program Design Data Sheet" (Martin, 1981);

Instrument #3, "Supplemental Data Sheet #1" (Martin, 1981); and

Instrument #4, "Supplemental Data Sheet #3" (Martin, 1981).

REFERENCES

Abert, J. G. (Ed.). (1979). Program evaluation at HEW: Research versus reality (in three parts). Part 2: Education. New York: Marcel Dekker.

Alkin, M. C. (1980). Naturalistic study of evaluation utilization. In L. A. Braskamp & R. D. Brown (Eds.), Utilization of evaluative information: New directions for program evaluation (pp. 19-28). San Francisco: Jossey-Bass.

Alkin, M. C., Kosecoff, J., Fitz-Gibbon, C., & Seligman, R. (1974). Evaluation and decision-making: The Title VII experience. Los Angeles: UCLA Center fro the Study of Evaluation.

Arias, M. B. (1979). Desegregation and the right of Hispanic students: The Los Angeles case. Los Angeles: UCLA Center for the Study of Evaluation.

Arias, M. B., Delgado, T., DePorcel, A., & Irzarry, R. (undated). Response to AIR study: Evaluation of the impact of ESEA Title VII special English bilingual education program. Unpublished report.

Baca, R. R. (1984). Bilingual education evaluation: An overview. Los Angeles: California State University, Evaluation Dissemination and Assessment Center.

Baker, E. L. (1980). Is something better than nothing? The metaphysical aspects of test design. In E. L. Baker & E. S. Quellmalz (Eds.), Educational testing and evaluation: Design, analysis, and policy (pp. 23-30). Beverly Hills: Sage.

Baker, J. R., Claus, R. N., & Manley, M. (1980). Meta-evaluation of the Sagiman Township Middle School Enrichment Center Project, 1979-80. Lansing: Michigan State Department of Education. (ERIC Document Reproduction Service No. Ed 206 720)

Baker, K. (1984, April). Ideological bias in bilingual education research. Paper presented at the meeting of the American Educational Research Association, New Orleans, LA.

Baker, K., & deKanter, A. A. (1981). Effectiveness of bilingual education: A review of the literature. Final draft report. Washington, DC: Office of Technical and Analytic Systems, U.S. Department of Education.

Baker, K., & deKanter, A. A. (1983). Federal policy and the effectiveness of bilingual education. In K. A. Baker & A. A. deKanter (Eds.), Bilingual education: A reappraisal of federal policy (pp. 35-36). Lexington, MA: Lexington Books, D. C. Heath.

Baker, K., & Pelavin, S. (1984). New directions in bilingual program evaluation. Paper presented at the meeting of the American Educational Research Association, New Orleans, LA.

Bank, A. (1980). An explosive change. Evaluation Comment, 6(2), 16.

Becker, H. A., Kirkhart, K. E., & Doss, D. (1982). Evaluating evaluation reports. Phi Delta Kappa CEVR Quarterly, 15(2), 18-20.

Bennett, W. J. (1985, November 22). Aide defends voucher and bilingual plans. Gainesville Sun, p. 2C.

Blumner, A. S. (1980). Evaluation demand and evaluation utilization. Evaluation Comment, 6(2), 12.

Boeckmann, M. E. (1981). Rethinking the results of a negative income tax experiment: A case of differential attrition. In R. F. Boruch, P. M. Wortman, & D. S. Cordray (Eds.), Reanalyzing program evaluations: Policies and practices for social and educational programs (pp. 341-356). Washington, DC: Jossey-Bass.

Boruch, R. F., Cordray, D. S., & Pion, G. M. (1981). How well are local evaluations carried out? In L. E. Datta (Ed.), Evaluation in change: Meeting new government needs (pp. 7-12). Beverly Hills: Sage.

Brandl, J. E. (1980). Policy evaluation and the work of legislatures. In L. A. Braskamp & R. D. Brown (Eds.), Utilization of evaluative information: New directions in program evaluation (pp. 37-43). San Francisco: Jossey-Bass.

Brannon, D. R. (1985). Toward excellence in secondary vocational education: Using evaluation results (Info. Series No. 294). Columbus: Ohio State University, National Center for Research in Vocational Education. (ERIC Document Reproduction Service No. ED 254 653)

Braskamp, L. A., & Brown, R. D. (Eds.). (1980). Utilization of evaluative information: New directions in program evaluation. San Francisco: Jossey-Bass.

Brown, R. D., & Braskamp, L. A. (1980). Summary: Common themes and a checklist. In L. A. Braskamp & R. D. Brown (Eds.), Utlization of evaluative information: New directions for program evaluation (pp. 91-97). San Francisco: Jossey-Bass.

Buros, O. K. (Ed.). (1972). The seventh mental measurements yearbook. Highland Park, NJ: Gryphon Press.

Burry, J. (1979). Evaluation in bilingual education. Evaluation Comment, 6(1), 1-14.

Burry, J. (1980). Toward a national research agenda for evaluation. Evaluation Comment, 6(2), 17-19.

Cahill, R. J., & Foley, J. J. (1979). Evaluation and evaluative research in an ur-an bilingual program. In J. G. Abert (Ed.), Program evaluation at HEW: Research versus reality. Part 2: Education (pp. 117-135). New York: Marcel Dekker.

Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research. Chicago: Rand McNally.

Campeau, P. L., Roberts, A. O. H., Bowers, J. E., Austin, M., & Roberts, S. J. (1975). The identification and description of exemplary bilingual education. Palo Alto, CA: American Institutes for Research.

Cardenas, J. A. (1977, June). AIR evaluation of bilingual education. Intercultural Development Research Association Newsletter, pp. 1-5.

Carlson, R. W. (1979). Pouring conceptual foundations: A utilization role and process for evaluation research. In H. C. Schulberg & J. M. Jerrell (Eds.), The evaluator and management (pp. 55-68). Beverly Hills: Sage.

Ciarlo, J. A. (Ed.). (1981). Utilizing evaluation: Concepts and measurement techniques (Vol. 6). Beverly Hills: Sage.

Cook, T. D. (1974). The potential and limitations of secondary evaluations. In M. W. Apple, M. J. S. Subkoviak, & H. S. Lufler (Eds.), Educational evaluation: Analysis and responsibility (pp. 115-122). Berkeley, CA: McCutchan.

Cook, T. D., & Gruder, C. L. (1978). Meta-evaluation research. Evaluation Quarterly, A Journal of Applied Social Research, 2(1), 3-52.

Danoff, M. N., Coles, G. J., McLaughlin, D. H., & Reynolds, D. J. (1977a). Evaluation of the impact of ESEA Title VII special/English bilingual education programs, Vol I: Study design and interim findings. Palo Alto, CA: American Institutes for Research.

Danoff, M. N., Coles, G. J., McLaughlin, D. H., & Reynolds, D. J. (1977b). Evaluation of the impact of ESEA Title VII special/English bilingual education programs, Vol. 2: Project descriptions. Palo Alto, CA: American Institutes for Research.

Danoff, M. N., Coles, G. J., McLaughlin, D. H., & Reynolds, D. J. (1978). Evaluation of the impact of ESEA Title VII special/ English bilingual education programs, Vol. 3: Year two impact data, educational process, and in-depth analyses. Palo Alto, CA: American Institutes for Research.

Datta, L. E. (1981a). Communicating evaluation results for policy decision making. In R. A. Berk (Ed.), Educational evaluation methodology: The state of the art (pp. 124-125). Baltimore, MD: The Johns Hopkins University Press.

Datta, L. E. (Ed.). (1981b). Evaluation in change: Meeting new government needs (Vol. 12). Beverly Hills: Sage.

DeGeorge, G. P. (1983). The guest editor speaks. Bilingual Journal 7(2), 6.

Development Associates. (1973). A process evaluation of the bilingual education program, Title VII, Elementary and secondary Education Act (Vol. 1). Washington, DC: U.S. Office of Education.

Dulay, H., & Burt, M. (1979a). The efficacy of bilingual education. Educational Evaluation and Policy Analysis, 1(5), 72-73.

Dulay, H., & Burt, M. (1979b). Research priorities in bilingual education. Educational Evaluation and Policy Analysis, 1(3), 39-53.

English, J. J. (1983, April). Talking points for problems and issues related to the meta-analysis of Title VII bilingual education project reports: A federal perspective. Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, Canada.

Evaluation Research Society Standards Committee. (1984). Evaluation Research Society standards for program evaluation. In R. F. Conner, D. G. Altman, & C. Jackson (Eds.), Evaluation Studies Review Annual, 9, 680-692.

Fuentes, E. J. (1986). OBEMLA creates research and evaluation staff under new Bilingual Education Act. FORUM, 9(3), 1.

Glaser, E. M. (1980). Strategies for obtaining utilizable knowledge. In L. A. Braskamp & R. D. Brown (Eds.), Utilization of evaluative information: New directions for program evaluation (pp. 83-90) San Francisco: Jossey-Bass.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). Meta-analysis in social research. Beverly HIlls: Sage.

Gowin, D. B. (1981). Philosophic analysis. In N. L. Smith (Ed.), New techniques for evaluation: New perspectives in evaluation (Vol. 2) (pp. 299-308). Beverly Hills: Sage.

Gowin, D. B., & Millman, J. (1978, March). Can meta-evaluation give a direction for research on evaluation? Paper presented at the 63rd Annual Meeting of the American Educational Research Association, Toronto, Ontario, Canada.

Gray, T. (1977). Response to AIR study: Evaluation of the impact of ESEA Title VII special/English bilingual education program. Arlington, VA: Center for Applied Linguistics.

Guba, E. G. (1975). Problems in utilizing the results of evaluation. Journal of Research and Development in Education, 8(3), 42-54.

Holley, F. M. (1980). Evaluation uses, methods, and measures. Evaluation Comment, 6(2), 12-13.

Joint Committee on Standards for Educational Evaluation. (1981). Standards for evaluations of educational programs, projects, and materials. New York: McGraw-HIll.

Kean, M. H. (1983). Administrative uses of research and evaluation information. In E. W. Gordon (Ed.), Review of research in education (Vol. 10) (pp. 361-415). Washington, DC: American Educational Research Association.

Lincoln, Y. S., & Guba, E. E. (1984, April). Research, evaluation, and policy analysis: Heuristics for disciplined inquiry. Paper presented at the Annual Meeting of the Evaluation Network, Evaluation Research Society, and the American Educational Research Association, New Orleans, LA.

Linn, R. L. (1981). A preliminary look at applicability of the educational evaluation standards. Educational Evaluation and Policy Analysis, 3(2), 87-91.

Lipsey, M. W., Crosse, S., Dunkle, J., Pollard, J., & Stobart, G. (1985). Evaluation: The state of the art and the sorry state of the science. In D. S. Cordray (Ed.), Utilizing research in evaluation planning: New directions for program evaluation (pp. 7-28). San Francisco: Jossey-Bass.

Littlejohn, J. M. (1981). Comments on final draft report of "The Effectiveness of Bilingual Education: A Review of the Literature" (Memorandum). Washington, DC: U.S. Department of Education, Office of Civil Rights.

Lopez, G., & Cervantes, R. (1978). 1978 abstract ofAIR report criticisms. Sacramento, CA: Department of Education.

Lyon, C., Dosher, L., McGranahan, P., & Williams. R. (1978). Evaluation and school districts (preliminary report). Los Angeles: UCLA, Center for Study of Evaluation.

Madaus, G. F., Scriven, M. S., & Stufflebeam, D. L. (Eds.), (1983). Evaluation models: Viewpoints on educational and human services evaluation. Boston: Kluwer-Nijhoff.

Maher, C. A. (1982). Utilization of program evaluation information in public schools: Perspectives and guidelines for school psychologists. Journal of School Psychology, 29(2), 113-121.

Martin, P. H. (1979). Research and bilingual program evaluation. San Francisco: California Evaluation Services.

Martin, P. H. (1980). Evaluation designs for student achievement data in bilingual education programs. San Francisco: California Evaluation Services.

Martin, P. H. (1981). Evaluation in bilingual education programs: A meta-evaluation of Title VII projects, Fiscal year 1979-80 (Doctoral dissertation, University of California at San Francisco, 1981). Dissertation Abstracts International, 43, 1115A.

Martin, P. H. (1982a). Considerations and recommendations for evaluating bilingual education programs. San Francisco: California Evaluation Services.

Martin, P. H. (1982b). Meta-analysis, meta-evaluation, and secondary analysis. San Francisco: California Evaluation Services.

Millman, J. (1981). A checklist procedure. In N. L. Smith (Ed.), New techniques for evaluation: New perspectives in evaluation (Vol. 2) (pp. 309-320). Beverly Hills: Sage.

Mitchell, J. V. (Ed.). (1985). Ninth mental measurements yearbook. Lincoln, NE: University of Nebraska, Buros Institute of Mental Measurements.

Mosteller, F., & Moynihan, D. P. (1972). On equality of educational opportunity. New York: Random House.

Mullarney, P. (1974). Evaluating community council effectiveness using Provus' discrepancy evaluation model. Community Education Journal, 4(3), 54-57.

Okata, M. (1983). Synthesis of reported evaluation and research evidence on the effectiveness of bilingual education basic projects, final report: Tasks 1-6. Los Alamitos, CA: National Center for Bilingual Research.

O'Malley, J. M. (1978, Winter). Review of evaluation of the impact of ESEA Title VII special/English bilingual education programs. Bilingual Resources, 1, 6-10.

Orata, P. T. (1940). Evaluating evaluation. Journal of Educational Research, 33(9), 641-661.

Popham, J. W. (1981). Crumbling conceptions of educational testing. In W. W. Welch (Ed.), Educational evaluation--Recent progress, future needs: Proceedings of the Minnesota Evaluation Conference, May 1980 (pp. 30-36). Minneapolis: Minnesota Research and Evaluation Center.

Provus, M. M. (1972). The discrepancy evaluation model. In P. A. Taylor & D. M. Cowley (Eds.), Readings in curriculum evalaution (pp. 117-127). Dubuque, IA: William C. Brown.

Rotberg, I. C. (1983). A reply to Baker. Harvard Educational Review, 53(1), 106.

Rutherford, W. L., & Hoffman, J. V. (1981). Toward implementation of the ESEA Title I evaluation and reporting system: A concerns analysis. Educational Evaluation and Policy Analysis, 3(4), 17-23.

Sanders, J. R. (1981). Reflections on evaluation plans for bilingual projects. Urbana: University of Illinois. (ERIC Document Reproduction Service No. ED 253 574)

Sanders, J. R., & Nafziger, L. J. (1976). Checklist for judging the adequacy of evaluation design. Urbana: University of Illinois.

Scriven, M. (1969). An introduction to meta-evaluation. Educational Product Report, 2(5), 36-38.

Scriven, M. (1974). Key evaluation checklist. Minneapolis: Minnesota Research and Evaluation Center.

Scriven, M. (1981). Evaluation: The state of the science. In W. W. Welch (Ed.), Evaluation studies review annual (Vol. 1) (pp. 119-139). Beverly HIlls:Sage.

Scriven, M. (1983). Evaluation ideologies. In G. F. Madaus, M. S. Scriven, & D. L. Stufflebeam (Eds.), Evaluation models: Viewpoints on educational and human services in evaluation (pp. 258-260). Boston: Kluwer-Nijhoff.

Scriven, M. (1985). Key evaluation checklist (rev.). Minneapolis: Minnesota Research and Evaluation Center.

Seidner, S. S. (1982). Political expedience or educational research: An analysis of Baker and deKanter's review of literature on bilingual education. In S. S. Seidner (Ed.), Issues of language assessment: Foundations and research (pp. 219-287). Springfield: Illinois State Board of Education.

Smith, N. L. (1981a). Creating alternative methods for educational evaluation. In N. L. Smith (Ed.), *New directions for program evaluation: Federal efforts to develop new evaluation methods* (No. 12) (pp. 77-94). San Francisco: Jossey-Bass.

Smith, N. L. (1981b). Drawing critical insights from new sources: The emerging field of meta-evalaution. In N. L. Smith (Ed.), *New techniques for evaluation: New perspectives in evaluation* (Vol. 2) (pp. 263-264). Beverly Hills: Sage.

Smith, N. L. (1981c). The progress of educational evaluation: Rounding the first bends in the river. In W. W. Welch (Ed.), *Educational evalaution--Recent progress, future needs: Proceedings of the Minnesota Evaluation Conference, May, 1980* (pp. 19-29). Minneapolis: Minnesota Research and Evaluation Center.

Stake, R. E. (1983). Program evaluation: Particularly responsive evaluation. In G. F. Madaus, M. S. Scriven, & D. L. Stufflebeam (Eds.), *Evaluation models: Viewpoints on educational and human services in evaluation* (pp. 303-305). Boston: Kluwer-Nijhoff.

Steinmetz, A. (1983). The discrepancy evalaution model. In G. F. Madaus, M. S. Scriven, & D. L. Stufflebeam (Eds.), *Evaluation models: Viewpoints on educational and human services in evaluation* (pp. 79-99). Boston: Kluwer-Nijhoff.

Stevenson, J. F., Longabaugh, R. H., & McNeill, D. N. (1979). Metaevaluation in the human services. In H. C. Schulberg & J. M. Jerrell (Eds.), *The evaluator and management* (pp. 37-54). Beverly Hills: Sage.

Stufflebeam, D. L. (1974a). *Meta-evaluation* (The Evaluation Center Occasional Paper Series #3). Kalamazoo: Western Michigan University, College of Education.

Stufflebeam, D. L. (1974b). *Toward a technology for evaluating evaluation*. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL.

Stufflebeam, D. L. (1978). Meta-evaluation: An overview. *Evaluation and the Health Professions*, *1*(1), 17-43.

Stufflebeam, D. L. (1981a). Metaevaluation: Concepts, standards, and uses. In R. A. Berk (Ed.), *Educational evaluation methodology: The state of the art* (pp. 146-163). Baltimore, MD: The Johns Hopkins University Press.

Stufflebeam, D. L. (1981b). Standards, research, and training: Three priorities for professionalizing educational evaluation. In W. W. Welch (Ed.), Educational evalaution--Recent progress, future needs: Proceedings of the Minnesota Evaluation Conference, May, 1980 (pp. 37-49). Minneapolis: Minnesota Research and Evaluation Center.

Stufflebeam, D. L., & Madaus, G. F. (1983). The standards for evaluation of educational programs, projects, and materials: A description and summary. In F. G. Madaus, M. S. Scriven, & D. L. Stufflebeam (Eds.), Evaluation models: Viewpoints on educational and human services in evaluation (pp. 395-405). Boston: Kluwer-Nijhoff.

Stufflebeam, D. L., & Shrinkfield, A. J. (1985). Systematic evaluation: A self-instructional guide to theory and practice. Boston: Kluwer-Nijhoff.

Tallmadge, G. K., Lam, T. C. M., & Camarena, M. L. (1985a). The evaluation of bilingual education programs for language-minority, limited-English-proficient students: A status report with recommendations for future development, Phase I, Executive summary (SRA Report no. 285). Washington, DC: U.S. Department of Education.

Tallmadge, G. K., Lam, T. C. M., & Camarena, M. L. (1985b). The evaluation of bilingual education programs for language-minority, limited-English-proficient students: A status report with recommendations for future development, Phase I, Technical recommendations document (SRA Report No. 285). Washington, DC: U.S. Department of Education.

Troike, R. C. (1978). Research evidence for the effectiveness of bilingual education. Washington, DC: National Clearinghouse for Bilingual Education.

United States Commission on Civil Rights. (1975). A better chance to learn: Bilingual-bicultural education (Clearinghouse Publication No. 51). Washington, DC: National Clearinghouse for Bilingual Education.

United States Department of Education. (1983, October). Application for grants under bilingual education program--Demonstration projects (CFDA Number 84.003D) (ED Form 4561). Washington, DC: Office of Bilingual Education and Minority Languages Affairs.

Vazquez, J. A. (1980). Evaluation needs in bilingual education. Evaluation Comment, 6(2), 11-12.

Weiss, C. H. (1972). Utilization of evaluation: Toward comparative study. In P. A. Taylor & D. M. Cowley (Eds.), Readings in curriculum evaluation (pp. 220-224). Dubuque, IA: William C. Brown.

Wholey, J. S., Scanlon, J. W., Duffy, H., Fukumoto, J. S., & Vogt, L. M. (1970). Federal evaluation policy: Analyzing the effects of public programs. Washington, DC: The Urban Institute.

Willig, A. C. (1982). The effectiveness of bilingual education: Review of a report. NABE Journal, 6(2-3), 1-19.

Willig, A. C. (1984). A meta-analysis of selected studies on the effectiveness of bilingual education (Doctoral dissertation, University of Illinois, Champaign, 1983). Dissertation Abstracts International, 43, 4515A.

Yates, J. R., & Ortiz, A. A. (1983, September). Baker & deKanter review: Inappropriate conclusions of the efficacy of bilingual education. NABE Journal, 7(3), 12-13.

Zappert, L. T., & Cruz, R. (1977). Bilingual education: An appraisal of empirical research. New York: Harper & Row.

BIOGRAPHICAL SKETCH

Bess Staes Fry spent childhood days in the little suburb of Montebello, California, just at the eastern boundary of the city of Los Angeles. She received her B.A. degree and teaching certificates from Whittier College, California, where her student teaching assignment sent her back to Montebello to the school she had attended 10 years before to intern with Mrs. Margaret Rowe, her own sixth grade teacher who had encouraged her to become a teacher. Over the years, the community had changed from predominately Anglo to Hispanic, but the schools had not kept pace with the changes in the language needs of the student population. A state mandated foreign language program for sixth graders started the young teacher on the long trek to the meta-evaluation of Title VII bilingual education projects. Key points in the journey included a National Defense Education Act Summer Language Institute, two weeks in Mexico, coronation as La Reina de la Fiesta de Navidad, and marriage to a Romance and Germanic Philologist and Medievalist, which started with a honeymoon year in Bucharest, Rumania, and ended 22 countries and 47 states later in Gainesville, Florida, as faculty wife, mother, widow, graduate student, and program evaluator.

Ms. Fry's professional career in the field of education includes 10 years of classroom teaching at the elementary, junior high school, and university levels, including teaching vice-principal of the American School, Bucharest, Rumania. She received her Master of Arts degree from the College of Arts and Sciences, University of Florida, in linguistics and reading, in 1978. She then continued her doctoral studies in the College of Education, University of Florida, where she was a Title VII Bilingual Fellow, majoring in curriculum and instruction and specializing in bilingual-multicultural education and program evaluation.

Ms. Fry has been an enthusiastic member of Phi Delta Kappa, holding numerous elected positions including President of the North Central Florida Chapter of Phi Delta Kappa, and six years as delegate to the 38th, 39th, and 40th Phi Delta Kappa International Biennial Councils, and the 1980, 1982, and 1984 Phi Delta Kappa District VII conferences. She also holds membership in the American Evaluation Association, the American Educational Research Association--Division H Evaluation, the National Association for Bilingual Education, and the American Society of Professional and Executive Women. In 1983, Ms. Fry started her own business, Educational Evaluation and Research Consulting Service, accepting numerous contracts for external evaluation from the state department of education and various school districts. Her publications are mainly in the field of evaluation.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Clemens L Hallman

Clemens L. Hallman, Chairman
Professor of Instruction and
Curriculum

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Sandra B. Damico

Sandra B. Damico
Professor of Foundations of
Education

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Robert G Wright

Robert G. Wright
Associate Professor of
Instruction and Curriculum

This dissertation was submitted to the Graduate Faculty of the College of Education and to the Graduate School, and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

December, 1986

David C. Smith (N)

Dean, College of Education

Dean, Graduate School